UNIVERSITY OF OKLAHOMA

GRADUATE COLLEGE

# SYSTEMS ANALYSIS OF GRASS CELL WALL BIOSYNTHESIS FOR BIOFUEL PRODUCTION

A DISSERTATION

SUBMITTED TO THE GRADUATE FACULTY

in partial fulfillment of the requirements for the

Degree of

DOCTOR OF PHILOSOPHY

By

FAN LIN
Norman, Oklahoma
2017

SYSTEMS ANALYSIS OF GRASS CELL WALL BIOSYNTHESIS FOR BIOFUEL PRODUCTION

A DISSERTATION APPROVED FOR THE
DEPARTMENT OF MICROBIOLOGY AND PLANT BIOLOGY

BY

_____
Dr. Laura E. Bartley, Chair

_____
Dr. Ben F. Holt III

_____
Dr. Scott D. Russell

_____
Dr. Marc Libault

_____
Dr. Randall S. Hewes

## Acknowledgements

I would like to express my deepest appreciation to my supervisor, Dr. Laura Bartley, for her guidance and advice throughout all projects. She is always encouraging and willing to help me with patience. By working with her, I learned more than anything I learned in a classroom. I would like to acknowledge my collaborators, their significant contributions to the data collection process and their useful discussion were essential for the completion of these projects. Especially thanks to Dr. Lobban, Dr. Mallinson and Dr. Waters for working together as an excellent interdisciplinary team on the thermal conversion project. I enjoy the useful meetings and discussions. I also would like to thank my committee members, Dr. Holt, Dr. Libault, Dr. Russell and Dr. Hewes. They spent their valuable time to meet with me and discuss about these projects. I also thank members of the Bartley lab, Russell lab, and Libault lab for useful discussions. Especially thanks to Dr. Zhao for the discussion on statistical analysis and network analysis. Last but not least, my deepest gratitude is to my parents, Wanshu Mu and Song Lin, I appreciate their sincere love and support. Projects in this dissertation were funded by U.S. Department of Energy, National Science Foundation, U.S. Department of Agriculture. Also, thank OU Student Travel Assistance Program for funding the travels to present these projects.

# Table of Contents

# List of Tables

# List of Figures

# Abstract

Grass cell walls play an important role in plant development, pathogen defense, and are an abundant and sustainable carbon source for biofuel production. Chapter 1 introduces the great impact of cell wall composition on the major biomass-based biofuel technologies, biochemical and thermal conversion. To further improve biomass conversion efficiency by manipulating cell walls genetically, it is important to understand the molecular basis of cell wall synthesis and regulation. However, current understanding of grass cell wall synthesis lags behind that of dicots, despite the great compositional difference between them. In this dissertation, I apply "omics" methods to further our knowledge of cell wall synthesis during grass development and demonstrate that genetic manipulation of cell walls can potentially improve thermal conversion.

In Chapter 2, we examine the cell wall composition and the transcripts of 65 putative cell wall synthesis genes in 30 rice samples from different organs at 10 developmental time points. A method is developed to identify candidate cell wall synthesis genes, based on the correlations between cell wall abundance and gene expression. We establish hypotheses for nine candidate genes that may synthesize cell wall components like xylans, mixed linkage glucan, and pectins. The cell wall profile also provides a baseline for evaluating the variation of cell wall composition.

In Chapter 3, we perform proteomics, cell wall profiling, and metabolite profiling on a rice elongating stem internode. With LC-MS/MS, we detect a total number of 2356 proteins in this internode. Many of them are glycosyltransferases, acyltransferases, glycosylhydrolases, cell wall-localized proteins, and protein kinases from families that may function in cell wall biosynthesis or remodeling or regulation. The presence of

these proteins is consistent with active cell wall synthesis in this internode, as indicated by cell wall assays. This study fills the void of shotgun proteomics data in rice stems and provides the basic information for a more detailed multi-omics experiment on internode segments.

Chapter 4 is a proof of principal study demonstrating that genetic manipulation of cell wall structures can improve the efficiency of thermal conversion. A major challenge for thermal conversion, thermal products with various chemical natures can not be upgraded to fuels with a simple catalytic strategy. One solution, known as thermal fractionation, is to collect different thermal products separately at different conversion temperatures, which depends on the different thermal stability of cell wall components. The efficiency of thermal fractionation can be improved by altering the thermal stability of cell wall components. We hypothesize that lignin thermal stability will increase in a switchgrass knock-down mutant of caffeic acid O-methyltransferase, which has less S-lignin subunits that can not form strong linkages at C-5 position. The elevation of lignin thermal stability will lead to a better segregation lignin-derived products from polysaccharide-derived products. Indeed, the results indicate that the mutant yields less lignin-derived products at a low conversion temperature, at which most polysaccharides are converted.

In conclusion, Chapter 2 and Chapter 3 develop an analysis method and a technical method for systems study of cell wall synthesis and identify a set of cell wall synthesis candidates for functional study. Understanding the function of more novel cell wall synthesis genes will provide more targets to manipulate cell wall structures. As a result,

there will be more opportunities to improve thermal fractionation efficiency other than the example we show in Chapter 4.

# Chapter 1: Background

Cell walls surround most plant cells and represent the most abundant source of organic carbon as the bulk of biomass (Gilbert 2010). The structure of plant cell walls is a scaffold of cellulose embedded in a cross-linked matrix consisting of lignin, hemicelluloses, pectins and other minor components such as cell wall proteins (Carpita 1996). Though similar in the basic structure, plant cell walls show a large diversity in composition in different species, in different cell types, and during plant development (Freshour et al. 1996; Martin et al. 2017). This diversity allows plant cell walls to serve a wide range of biological functions like maintaining cell shape, controlling growth rate, mechanical support, pathogen defense and source of signaling molecules and development cues (Keegstra 2010; Sørensen et al. 2010; Wolf et al. 2012). Besides biological functions, cell walls also have great industrial importance since they directly influence the quality of wood, forage, and dietary fiber in foods (Barnett and Bonham 2004; Jung and Allen 1995; Waldron et al. 2003). In recently years, biofuels made from biomass as a replacement of fossil fuels are receiving more attention due to global warming (Himmel and Bayer 2009). Biomass-biofuel conversion is directly impacted by cell walls since they represent up to 90% of biomass (Vogel 2008).

## Biological Functions of Cell Walls

Cell walls form an extracellular matrix that enclose plant cells and directly affect plant growth (Wolf et al. 2012). Since cell walls and the extracellular environment usually contain fewer solutes than cytoplasm, so that turgor pressure that pushes outward against cell walls is created, which is the main driving force for cell expansion. The

balance between turgor pressure and extensibility of cell walls determines plant cell growth (Lamond 2002). During cell growth, the surrounding cell walls are flexible primary walls. The high extensibility of primary walls allows cells to expand 10- to more than 1,000-fold between their initial formation at cell division until cessation of growth (Cosgrove 2016a). At the late stage of cell growth, a more rigid secondary cell wall is deposited in many cell types during the slow-down of cell growth. Wall extensibility during development is mainly influenced by crosslinks among cell wall polymers and by cell wall remodeling events. For example, methyl-esterified pectins can be de-esterified to form crosslinks among molecules at the end of cell elongation (Goldberg et al. 1986). Another example is the deceleration of leaf elongation in tall fescue associated with the accumulation of diferulate cross-links (MacAdam and Grabber 2002). Many cell wall remodeling enzymes play a crucial role in adjusting cell wall extensibility. For example, the hydrogen bonds between cellulose and xyloglucan, a hemicellulose, can be disrupted by a group proteins called expansin to promote cell wall loosening (Cosgrove 2005). In another example, enzymes from the glycohydrolase 9 family (GH9) possess cellulose activity *in vitro* and *in vivo*, which suggests they may loosen the cell wall to facilitate cell growth (Ohmiya et al. 1995; Park et al. 2003; Yoshida and Komae 2006).

Cell walls also determine the direction of anisotropic cell growth (Baskin 2005). During the growth of many cells in root and stem, cellulose microfibrils are wrapped around cells in a largely transverse orientation (Bashline et al. 2014a; Lamond 2002). This directs cell elongation in the longitudinal direction (along the axis of stem or root),

perpendicular to the orientation of cellulose microfibrils. At the end of elongation, cellulose microfibrils will be deposited in both longitudinal and oblique orientations to slow cell elongation in the longitudinal direction (Bashline et al. 2014a) .

Another basic function of cell walls is to provide mechanical support. The tension created by cell walls and turgor pressure provides mechanical rigidity for plant tissues and helps a plant to remain erect. This tension is also important for organ morphogenesis at the shoot apical meristem (Beauzamy et al. 2015). For specialized cells, such as sclerenchyma and fiber cells, mechanical strength is mainly from their secondary cell walls, which can be as thick as 2–6 μm (Haigler et al. 2012; Muraille et al. 2017). A "Bioinspired lignocellulosic film" experiment indicates that cellulose could be the main contributor of mechanical property in secondary cell walls but other cell wall components like lignin and xylan also play a role (Muraille et al. 2017).

Another important biological function of the cell wall is pathogen defense (Zhao and Dixon 2014). As the outmost physical barrier of cells, most prospective microbial pathogens have to interact with cell walls at the beginning of infection. The damaged cell wall will generate damage-associated molecules, which can bind to cell wall integrity sensors on plasma membrane such as receptor-like kinases to trigger downstream defense related signaling pathways (Boller and Felix 2009; Malinovsky et al. 2014). For example, the oligogalacturonides (OGs) released from pectins can activate WAK1, a wall-associated kinase, and downstream defense responses against fungal and bacterial pathogens in Arabidopsis (Brutus et al. 2010). As one of the

defense response, plants will remodel and reinforce the cell wall at the site of infection to limit the access of pathogens (Dong et al. 2008). Phenolic polymers like lignin, which can not be easily degraded by pathogens are deposited to cell walls to prevent further invasion of pathogens (Underwood 2012). While the mutation of a lignin synthesis gene *CAD*, can cause an increase in susceptibility to pathogens (Tronchet et al. 2010), though effects of cell wall biosynthesis mutants on defense are mixed.

**Cellulosic Biofuel**

The "Billion-Ton Vision" has reported that the United States has the potential to produce at least one billion dry tons of biomass resources on an annual basis without direct negative environmental impacts if under reasonable land usage policy (Tilman et al. 2009; US-DOE 2005). This amount of biomass could be used to produce enough bioenergy to displace about 30% of 2005 U.S. petroleum consumption without reducing food production (US-DOE 2016). Many benefits can be achieved by replacing petroleum with bioenergy such as ensuring future energy security, lowering greenhouse gas emission, producing less toxic byproducts, and improving the U.S. trade balance (US-DOE 2016). Two relatively well-developed biofuel technologies are biochemical conversion and thermal conversion. For biochemical conversion, biomass is deconstructed by enzymes into monosaccharides and fermented to mainly, ethanol, by micro-organisms. For thermal conversions, particularly pyrolysis and torrefaction, biomass is heated in a non-oxygen environment and the liquid products are collected for downstream chemical upgrade through catalytic reactions.

In the last decade, great advancement has been achieved in developing biochemical conversion technologies and applying them to establishing the lignocellulosic biofuel industry. Several bioconversion facilities have been initiated at the commercial scale including Beta Renewables, Fiberight, INEOS, Abengoa, American Process, DuPont, POET-DSM Advanced Biofuels, and GranBio. The total capacity of these facilities is about 90 millions gallons (US-DOE 2015). However, the current capacity is still far from the target of 21 billion gallons of renewable fuels derived from non-cornstarch feedstock by 2022 under the Energy Independence and Security Act of 2007. To accelerate the expansion of the biofuel industry, there is continuing need to improve conversion methods and the quality and quantity of feedstocks to support the U.S. Government goal of 30% production of transportation fuel from alternative sources by 2030. The successful scale-up of lignocellulosic biofuel production requires biomass feedstocks to be produced, processed, and transported efficiently.

Yield of biomass is especially important in terms of sustainability. As the world population rises to an estimated 9.3 billion by the year of 2050, there will be an increasing demand for energy and foods (Warnasooriya and Brutnell 2014). To achieve more efficient land usage and cost reduction, improving biomass yield per acre is important. Biomass yield can be improved by breeding or engineering to enhance light capture efficiency and prolong the biomass accumulation period. Light capture efficiency become a limitating factor for yield especially when crops are grown in high density (Warnasooriya and Brutnell 2014). Leaf architecture such as leaf angel and leaf orientation, and the chlorophyll density in upper and lower leaves are all good targets

for improving light capture efficiency (Djakovic-Petrovic et al. 2007; Ort et al. 2011; Sinclair and Sheehy 1999; Zhao et al. 2013). Since leaves and stems represent the bulk of biomass, increasing the time of vegetative growth by postponing flowering will improve overall biomass yield. For example, over expression of the floral repressor gene FLC in tobacco delays flowering and increases biomass yield (Salehi et al. 2005).

Though yield is important, biomass has to be efficiently converted to fuels. The bottleneck for biochemical conversion is deconstruction of plant cell walls into lignocellulosic-derived sugars. The sugar yield is mainly limited by the low deconstruction efficiency of cell wall, also known as, recalcitrance. Without pretreatment, the recovery efficiency of lignocellulosic-derived sugars for biochemical conversion is about 10-20% (Fu et al. 2011). With moderate routine pretreatment, the efficiency can be more than 60% though extra chemical or mechanical inputs are required (Fu et al. 2011; Fu et al. 2012; Willis et al. 2016). Many cell wall factors such as lignin, cellulose crystallinity, hemicelluloses, pectins, cell wall proteins, and hydroxycinnamic acids as well as the interactions among polymers affect cell wall recalcitrance (Tavares et al. 2015). Cell wall recalcitrance can be decreased by chemical pretreatments (Alonso-Simon et al. 2010; Tavares et al. 2015) and by genetic engineering to manipulate the structure and content of cell walls (Chen et al. 2013; Fu et al. 2011; Wilkerson et al. 2014).

Transportation and logistics are also constraints for the scale-up of biorefineries (Richard 2010). The total cost of transportation increases as facility capacity increases since the size of the facility determines the biomass draw area (Sultana et al. 2010). At a

scale of 200 to 1000 megaliters, the cost of biomass transport can be greater than the savings of larger biorefinery equipment (Aden A. 2002). Therefore, it is important to reduce transportation cost by densifying biomass, such as by creating pellets and cubes. Densification can be coupled with pretreatment processes that prepare the feedstock for downstream conversion. For biochemical conversion technologies, ammonia fiber expansion pretreatment can be applied to pelletize biomass and increase digestibility at the same time (Bals et al. 2010). For thermal conversion, the biomass can be converted on site into liquid fuel precursors that are more dense (Uslu et al. 2008). Mobile thermal conversion devices are under development and the current bench-scale machine developed by Battelle Corp. has a throughput of 50 pounds per day.

For thermal conversion, the biggest challenge is heterogeneity of thermal products that prevents efficient downstream upgrade processing. Though many catalytic strategies have been developed, the catalysts usually perform well on one class of thermal products but poorly on other products or even inhibited by the presence of other products (Carlson et al. 2008; Lin et al. 2015; Rezaei et al. 2016). To improve catalyst performance, one separation approach called thermal fractionation has been developed to separate the thermal products during conversion (Herron et al. 2017). This approach uses a successively increasing series of temperatures to fractionate the biomass into different product streams. Since the basis of thermal fractionation is the different thermal stability of cell wall components, more efficient thermal fractionation could be achieved by altering chemical bonds in cell walls (Herron et al. 2017).

Grass biomass is a promising feedstock for biofuel production. Cultivated grasses (family Poaceae) such as rice and maize are the most abundant, sustainable class of terrestrial biomass that can be produced in the U.S. (~57%) (US-DOE 2011). Globally, rice straw alone constitutes 23% of world agricultural waste (Lal 2005), a major potential source of biomass for fuel production. As a leading grass energy crop, switchgrass has high biomass yield, minimal water and nutritional requirements, widespread adaptability and high net energy gain (Baxter et al. 2014). Due to its perennial nature, switchgrass also have less cultivation input and higher nutrient use efficiencies than annual energy crops such as sorghum (van der Weijde et al. 2013). However, because switchgrass is an outcrossing tetraploid, interpretation of gene expression is challenging due to the complexity of multiple alleles. Furthermore, though now routine (Li and Qu 2011), switchgrass transformation is slow and reverse genetic resources are not available, hindering model testing. Thus, we have to first study grass cell wall biosynthesis in model grass organisms such as rice and then applied the knowledge to manipulate switchgrass biomass to enhance biofuel production.

Switchgrass is highly diverse in growth habit and drought tolerance, and can be divided into two ecotypes. Upland plants typically have better drought tolerance and produce more tillers than lowland plants. Lowland plants (usually tetraploid) are higher and grow better under flooded conditions than upland plants (Porter 1966). Relative to upland cultivars under the same conditions, lowlands tend to have a longer growing season and longer, thicker stems, contributing to lowlands typically accumulating

greater biomass than uplands (Wullschleger et al. 2010). However, most lowlands are cold-sensitive and fail to overwinter in northern, continental locations.

**Grass Cell Wall Composition and Biosynthesis**

Typical angiosperm plant cell walls consist of cellulose, hemicelluloses, lignin, pectins, proteins, and other minor components. Though the composition of cell walls varies significantly among species, cell types and developmental stages, cellulose, hemicellulose, and lignin are usually the three major components by the percent dry weight (Pauly and Keegstra 2010). Cellulose is a linear homogenous polysaccharide composed of glucose. Hemicelluloses encompass various heterogeneous polysaccharides with different monosaccharide composition and sidechain substituents (Scheller and Ulvskov 2010). Lignin is a large phenolic polymer composed of three major subunits: guaiacyl (G), syringyl (S), and p-hydroxyphenyl (H) unit (Vanholme et al. 2010a). The subunits can form various covalent bonds across each other and also be decorated by HCAs, tricin and other substituents (Lan et al. 2015; Vanholme et al. 2010a). Due to the complexity, the structure of natural lignin is still not fully understood.

Grass cell walls are different from dicot cell walls in terms of hemicelluloses, pectins, and cross-links among cell wall components (Vogel 2008). The most abundant hemicellulose in grass is the arabinoxylan, while xyloglucan is most abundant one in dicots cell walls. Grass xylans are usually substituted by considerable amount of arabinose residues and hydroxycinnamic acids (HCAs), namely, *para*-coumaric acid

(*p*CA) and ferulic acid (FA) (Rennie and Scheller 2014). Dicot xylans have more glucuronic acid substituents but much less arabinose substituents and barely any HCAs (Doering et al. 2012). Since dicot xylans almost contain no ferulic acid, the ferulic dimers that cross-link xylans and lignin become a unique feature of grass cell walls. The second most abundant grass hemicellulose, mixed-linkage glucan, is absence in dicots and other monocots expect horsetails (Xue and Fry 2012). Mainly present in primary cell walls, pectins represent up to 25% of primary cell walls in dicots but only 5% in grasses. The variation of cell wall composition among grasses is not as dramatic as the difference between grass and dicots (Pauly and Keegstra 2008). For example, the cellulose content in switchgrass and rice are both about 32%. The lignin and hemicellulose content in switchgrass is about 17% and 26%, slightly higher than the 12% and 24% in rice straw, respectively (Kimball Christensen 2008).

The complex structure of various cell wall polymers requires a large set of genes for synthesis and modification. More than 2000 genes are estimated to be required for metabolism of cell wall components (McCann and Rose, 2010). Currently, most genes in the phenylpropanoid pathway, which synthesizes lignin subunits, are well studied in Arabidopsis by genetic approaches (Fraser and Chapple 2011) and some of them were also characterized in rice. Genes in the phenylpropanoid pathway have relative conserve function across species, for example, the COMT mutant show similar phenotypes in Arabidopsis, tobacco, Alfalfa, sorghum, and switchgrass (Fornale et al. 2017; Fu et al. 2011; Goujon et al. 2003; Guo et al. 2001; Sattler et al. 2014). Cell wall polysaccharides are synthesized by several families of glycosyltransferases (GTs) and modified by

acyltransferases (ATs) from a grass-expanded clade of the BAHD acyltransferases family (Mitchell et al. 2007). Cellulose synthases from the GT2 family are known to form complexes that synthesize cellulose microfibrils at the plasma membrane (Bashline et al. 2014b). For the biosynthesis of arabinoxylans, the main grass hemicellulose, enzymes from the GT43, GT47, GT61, and GT75 families play a role (Rennie and Scheller 2014). These GT families consist of 15-41 members in rice (Cao et al. 2008) but only a few of them have been functionally characterized (Chen et al. 2013; Chiniquy et al. 2012; Konishi et al. 2007). The grass unique hemicellulose, mixed linkage glucan (MLG), is synthesized primarily by CSLF8, CSLH1, and CSLF6 from GT2 family (Burton et al. 2006; Vega-Sanchez et al. 2012). In recently years, genetic studies have revealed ATs are the enzymes that incorporate *p*CA to arabinoxylans or incorporate FA to lignin (Bartley et al. 2013; Petrik et al. 2014; Sibout et al. 2016). However, there are remaining questions about the grass cell wall synthesis like which gene incorporates FA to xylans and how does xylan backbone synthesis initiate (Faik 2010).

These cell wall synthesis families all contain multiple members in grasses. Cell wall-related GT families in rice such as GT61 can up to 40 members (Cao et al. 2008). The functions of these members may be redundant or non-redundant. For example, a study on four genes in GT43 family have shown they belong to 2 functionally nonredundant groups, but may have functional redundancy within groups (Lee et al. 2014b). The families of phenylpropanoid enzymes also have four to ten members in rice. For example, 4CL families have 5 members, and biochemical characterization of these 4CL

genes shown some differences of substrate specificities and turnover rates (Gui et al. 2011)

The regulation of cell wall biosynthesis has been most intensively studied in Arabidopsis but also in rice. About 30 transcription factors from six protein families have been studied in Arabidopsis with genetic approaches and yeast one-hybrid experiment (Handakumbura and Hazen 2012; Taylor-Teeples et al. 2015; Zhong et al. 2008). However, the transcriptional regulatory hierarchy of the entire pathway is still under investigation (Taylor-Teeples et al. 2015). Protein kinases also play a role in regulating cell walls synthesis since several protein kinase mutants show cell wall or vascularization phenotypes (Hematy et al. 2007; Matschi et al. 2013; Oh et al. 2011). For example, the overexpression of THE1 causes accumulation of ectopic lignin in seedlings (Hematy et al. 2007). However, which phosphorylation positions and how they connect to the cell wall regulatory network remains enigmatic. In addition, a recent study showed that the stunted growth of a lignin-deficient Arabidopsis mutant can be rescued by the disruption of a Mediator complex (Bonawitz et al. 2014). This result indicates that two Mediators are in an transcriptional pathway that links lignin biosynthesis to plant growth but the genes upstream and downstream of Mediators are still under investigation (Mach 2015). Knowledge of grass cell wall biosynthesis regulation lags behind that for Arabidopsis (Handakumbura and Hazen 2012) though some evidence has indicated the transcription factors regulating major cell wall components possess conserved in functions (Shen et al. 2012; Zhao and Bartley 2014; Zhong et al. 2011).

Given there are still gaps in our knowledge about cell wall synthesis and regulation, especially in grasses, it is important to identify other players by systematic studies. The dynamic changes of cell wall synthesis during development facilitate identification of cell wall related genes by transcriptomics (Shen et al. 2013; Zhang et al. 2014). Combining -omics data with other large-scale measurements such as cell wall profiling can be a powerful systematic approach that allows more direct result interpretation (Guo et al. 2014; Redestig and Costa 2011). Though they will require proper experiment design and new quantitative analysis methods, combining multiple -omics techniques such as transcriptomics, proteomics, and phosphoproteomics will allow identification of posttranscriptional regulation events and increase predictability of gene regulation network (Walley et al. 2016). The development of more comprehensive systematic approaches will allow us complete the full picture about cell wall synthesis and regulation, which will provide more opportunity and strategies for genetic manipulation. In addition, understanding how regulation of cell wall synthesis are coupled to the regulation of other biological processes such as plant growth will help reduce the potential defects in genetically manipulated energy crops.

As the understanding of cell wall biosynthesis grows, another important question is how this knowledge can be applied to improve the efficiency of biomass conversion. For biochemical conversion, genetic modifications on key recalcitrance factors such as lignin and FA cross-links can significantly increase sugar yield (Bartley et al. 2013; Loque et al. 2015). However, most of these studies are either based on model organisms or energy crops growing in lab conditions. There has been considerable speculation about how cell wall altered feedstocks would perform in the field when they exposed to

pests, pathogens, and temperature changes (Pedersen et al., 2005). In addition, significant differences in cell wall composition between rice cultivars growing in the greenhouse and field has been reported (Tanger et al. 2015). Therefore, it is important to evaluate if beneficial traits from small scale greenhouse experiments are still significant in energy crops grown in field. One successful example is a two-year field study on field-grown switchgrass COMT mutants showing consistently higher sugar yield than the wild type biomass (Baxter et al. 2014).

For thermal fractionation, improved thermal product segregation efficiency is the primary goal. The decomposition temperature range of lignin is wide and partially overlapped with decomposition temperature range of cellulose and hemicellulose. Therefore, though some parts of lignin are converted at a higher temperature than cellulose, the segregation of lignin-derived products from cellulose derived products is less efficient. In theory, modifying lignin structure for higher thermal stability will raise the decomposition temperature of lignin and facilitate better product segregation. Though this is a promising direction, no genetically manipulated biomass has been tested for this purpose so far.

In this dissertation, I have studied grass cell wall biosynthesis at the system level and demonstrated that genetically manipulated biomass can improve thermal fractionation efficiency. In Chapter 2, we analyzed the atlas of gene expression and cell wall composition in 30 rice samples of different tissues and developmental stages. Candidate grass cell wall synthesis genes were identified by the correlations between gene expression and abundance of cell wall components. Chapter 3 constitutes a proteomics study on an elongating rice stem internode where cell wall synthesis is active. The

presence of several uncharacterized GTs, ATs and regulators suggest they may have a function in stem cell wall development. In Chapter 4, we did thermal conversion and analyzed the thermal products on field-grown switchgrass COMT mutants. The results demonstrate more efficient thermal fractionation of lignin-derived products in mutants than in the wild-type plants.

# Chapter 2: Cell Wall Composition and Candidate Biosynthesis Gene Expression During Rice Development

**Authors:** Fan Lin, Chithra Manisseri, Alexandra Fagerström, Matthew L. Peck, Miguel E. Vega-Sánchez, Brian Williams, Dawn M. Chiniquy, Prasenjit Saha, Sivakumar Pattathil, Brian Conlin, Lan Zhu, Michael G. Hahn, William G.T. Willats, Henrik V. Scheller, Pamela C. Ronald, Laura E. Bartley

**Publication Status**: This chapter has been published in Plant and Cell Physiology.

**Author contributions:** F.L. and L.E.B. analyzed the data, interpreted the results and wrote the manuscript. F.L., L.E.B., M.E.V.-S., P.C.R. and H.V.S. revised the manuscript. L.E.B., P.C.R., and H.V.S. designed the study. L.E.B., D.M.C., and B.W. collected the samples. C.M., M.L.P., M.E.V.-S., L.E.B., and H.V.S. executed the cell wall chemical assays. L.E.B., B.W., B.C., P.S., and P.C.R. executed the qPCR. A.F. and W.G.T.W. executed the CoMPP assay. S.P. and M.G.H. executed the glycome profiling. L.Z. contributed to the statistical analyses and interpretations. All authors read and approved the final version of the manuscript.

**Abstract**

Cell walls of grasses, including cereal crops and biofuel grasses, compose the majority of plant biomass and intimately influence plant growth, development, and physiology. However, the functions of many cell wall synthesis genes, and the relationships among and the functions of cell wall components remain obscure. To better understand the patterns of cell wall accumulation and identify genes that act in grass cell wall biosynthesis, we characterized 30 samples from aerial organs of rice (*Oryza sativa* cv. Kitaake) at 10 developmental time points，3 to 100 days post germination. Within these samples, we measured 15 cell wall chemical components, enzymatic digestibility, and 18 cell wall polysaccharide epitopes/ligands. We also used quantitative reverse transcriptase PCR to measure expression of 50 glycosyltransferases, 15 acyltransferases and 8 phenylpropanoid genes, many of which had previously been identified as being highly expressed in rice. Most cell wall components vary significantly across development and correlations among them support current understanding of cell walls. We identified 92 significant correlations between cell wall components and gene expression and establish nine strong hypotheses for genes that synthesize xylans, MLG, and pectin components. This work provides an extensive analysis of cell wall composition throughout rice development, identifies genes likely to synthesize grass cell walls, and provides a framework for development of genetically improved grasses for use in lignocellulosic biofuel production and agriculture.

## Introduction

Cell walls play a key role in plant vegetative development. They provide mechanical support for plant aerial structures (Carpita 1996) and determine the morphology and extension of cells (Cosgrove 2005; Ryden et al. 2003; Sasayama et al. 2011). Changes in cell wall composition alter plant organ biophysical properties that contribute to lodging resistance (Wang et al. 2012) and pathogen resistance (Zhao and Bartley 2014), with direct implications for crop production. In addition, cell walls in biomass are a source of animal feed and have potential to be converted to transportation fuels. In particular, stems and leaves of cereal crops, such as maize (*Zea mays*), rice (*Oryza sativa*), wheat (*Triticum aestivum*), sorghum (*Sorghum bicolor*), and related bioenergy grasses, like switchgrass (*Panicum virgatum*), represent over half of the estimated billion tons of biomass that can be sustainably produced in the U.S. for lignocellulosic biofuels (US-DOE 2011). Furthermore, rice straw alone, comprises almost one quarter of all agricultural waste globally (Lal 2005). Recalcitrance to enzymatic digestion remains a major bottleneck for efficient biochemical conversion of lignocellulosic material to biofuels (Bartley et al. 2013; Lynd et al. 2008). Cell wall biosynthesis and modification has been most extensively studied in the dicotyledenous reference plant, *Arabidopsis thaliana*. However, much remains to be learned about grass cell walls, which differ from those of dicots in terms of composition and patterning. For example, many questions remain as to which proteins function in grass cell wall synthesis and the contributions of grass cell wall components to cell wall strength, flexibility and plant growth have not been well determined. To reduce this gap, we present here a study of

grass cell wall content across development and its relationship to expression of putative cell wall biosynthesis genes.

Structural components of plant cell walls include cellulose microfibrils, matrix polysaccharides, lignin, minerals, and structural proteins. Primary cell walls form at the cell plate during division and add to expanding cells; whereas, secondary cell walls form around some cells after cessation of growth. Compared to those of eudicots, grass cell walls differ in the abundance and structures of matrix polysaccharides (i.e., hemicelluloses and pectin) and proteins; modifications on lignin; and silica (Ishii 1997; Vogel 2008). Grass primary walls contain more glucuronoarabinoxylans (GAX) and mixed linkage glucans (MLG); but less xyloglucan (XyG), pectins, mannans, and structural proteins (Fincher 2009; Scheller and Ulvskov 2010; Vogel 2008).

Xylan is the major grass hemicellulose. It consists of a β-(1→4)-linked xylose (Xyl) backbone with α-(1→2)- or α-(1→3)-linked arabinose (Ara) moieties. To a minor extent compared with dicots, the xylose backbone can also be substituted by α-(1→2)-linked glucuronic acid (GlcA) and 4-*O*-methyl-GlcA (Rennie and Scheller 2014; Scheller and Ulvskov 2010; Zeng et al. 2010). Some of the α-(1→3)-linked Ara residues are esterified with hydroxycinnamic acids (HCAs, primarily, ferulic acid [FA] and *p*-coumaric acid [*p*CA]) at the *O*-5 position. The ferulic acid on GAX can undergo radical-oxygen mediated cross-linking to form dehydro-ferulate dimers, as well as higher-order oligomers and ether bonds with lignin (Buanafina 2009). A variety of minor side chains have also been documented on grass xylan such as β-D-Xyl-(1→2)-α-L-Ara-(1→3)- and β-D-Gal-(1→4)-β-D-Xyl-(1→2)-α-L-Ara-(1→3)- (Chiniquy et al.

2012; Pena et al. 2016). However, the structure of grass GAX is still not fully understood. For example, it is unknown if the tetrasaccharide, 4-β-d-Xyl-(1–4)-β-d-Xyl-(1–3)-α-l-Rha-(1–2)-α-d-GalA-(1–4)-d-Xyl, at reducing ends of dicot xylan also exists in grass GAX. This reducing end tetrasaccharide may be an initiator or terminator for dicot xylan backbone synthesis (York and O'Neill 2008), which raises the question of how the synthesis of grass GAX initiates or terminates.

Other components that are different in grasses compared to dicots include MLG, an unbranched glucan with both β-(1→4)- and β-(1→3)-linkages, which has been only found in grasses, horsetail (*Equisetum*), and some ferns (Xue and Fry 2012). Though present, pectins, arabinogalactan proteins (AGPs), and other structural proteins have low abundance and have received relatively little attention in grasses (Atmodjo et al. 2013; Carpita 1996). Grass lignin is esterified by HCAs, primarily *p*CA, and etherified by the flavone tricin (Lan et al. 2015; Ralph 2010). In addition, acetylation of grass lignin predominantly occurs on guaiacyl rather than syringyl units, the opposite of dicots (del Rio et al. 2012). Finally, grass cell walls generally accumulate more silicon, mainly in forms of silica ($SiO_2 \bullet nH_2O$) (Carnelli et al. 2001; Ma and Yamaji 2006; Parry et al. 1984).

Though incomplete, progress has been made in revealing the proteins that function in grass-diverged aspects of cell wall synthesis. Glycosyltransferases (GTs) are an enzyme superfamily with 97 families, many of which form the glycosidic bonds of cell wall polysaccharides (Cosgrove 2005; Scheible and Pauly 2004). In grasses, CSLFs and

CSLHs from the GT2 family function in the synthesis of MLG (Burton and Fincher 2012; Burton et al. 2008; Burton et al. 2006; Doblin et al. 2009). Mutations in several GT43s and GT47s are known to reduce the synthesis of the GAX backbone (Chen et al. 2013; Lee et al. 2014a). Why multiple GTs appear to be utilized for the synthesis of xylan backbone is still unknown, since a GT47 (IRX10) is capable of synthetizing a backbone structure in vitro in the absence of other proteins (Jensen et al. 2014; Urbanowicz et al. 2014). Some GT61s add side groups to xylan that are unique to or more abundant in grasses (Anders et al. 2012; Chiniquy et al. 2012). Acyl-CoA acyltransferases (ATs) from a particular subclade of the so-called "BAHD" family modify grass cell wall precursors. Specifically, OsAT10, incorporate *p*CA into GAX, (Bartley et al. 2013; Withers et al. 2012). However, an acyltransferase that incorporates FA into GAX and other polysaccharides have still not been unambiguously identified (de Oliveira et al. 2015), though AT1 has been implicated (Buanafina et al. 2016).

Several reverse genetics studies of cell wall synthesis have been directed by the hypothesis that grass-diverged cell wall synthesis genes are more highly expressed relative to their homologs in dicots (Mitchell et al. 2007). Such clades often possess more members in grasses (i.e., are grass-expanded) relative to other groups of putative cell wall biosynthesis genes (Bartley et al. 2013; Cao et al. 2008; Penning et al. 2009), as is the case for the GT61 and the cell wall AT genes. Moreover, the synthesis enzymes and functions of other less abundant components such as pectins and the oligosaccharides on AGPs have mainly been studied in Arabidopsis (Atmodjo et al. 2013; Harholt et al. 2010; Mohnen 2008). Even when polymers are very similar

between dicots and grasses, the problem remains of which genes in large protein clades possess a given function in a given cell type, organ or developmental time point. Identifying the genes that function in grass wall polymer biosynthesis as well as understanding the patterns of accumulation will allow us to use genetic means to examine the functions of different cell wall components in grass biology and gain insight into the fundamental question of the selective basis of cell wall evolution.

Rice is an excellent model to study grass cell wall synthesis due to its well-annotated genome, abundant genetic resources (Krishnan et al. 2009; Li et al. 2016), ease of genetic transformation, and well-described development (Itoh et al. 2005; Sylvester et al. 2001). As with other grasses, changes in rice cell wall composition accompany development of organs and passage along the stages of plant development, e.g., seedling, vegetative, reproductive (Azuma et al. 1996; Guo et al. 2014; Rancour et al. 2012). Though distinct morphological and chemical differences have been observed between the juvenile and adult vegetative stages in other grasses, such as maize (Abedon et al. 2006; Williams et al. 2000), the development of rice is continuous and there is not a distinct juvenile to adult boundary (Itoh et al. 2005; Sylvester et al. 2001). The leaves and internodes of grasses mature from tip to base and from outer or basal organs to inner, distal organs. Hence, newly formed organs and basal parts of organs nearest the meristems possess young, still growing cell walls, while older organs and distal organ parts have ceased both primary growth and secondary wall formation.

Changes in metabolite abundance and gene expression during plant and organ development or under genetic perturbation have allowed researchers to study correlations among metabolites, genes, and traits (Christensen et al. 2010; Guo et al. 2014; Taboada et al. 2010; Yonekura-Sakakibara et al. 2008). Small-scale correlation analysis has been used to study lignin components that reduce silage digestibility (Jung and Allen 1995; Taboada et al. 2010) and secondary cell wall components that alter enzymatic digestibility (Li et al. 2015; Van Acker et al. 2013). Recently, correlation analysis based on microarray data and cell wall composition data of rice identified correlations between lignin and major monosaccharide components with gene modules composed of hundreds of genes (Guo et al. 2014). That cell walls are abundant, structural components may facilitate correlation analysis (Rajasundaram et al. 2014). On the other hand, the fact that most cell wall components co-occur may confound correlation analyses, requiring the use of other information, such as knowledge of associations among cell wall components and phylogenetic relationships among putative synthesis proteins.

Here, we report the comprehensive measurement of cell wall chemical and polymer composition, and expression of putative cell wall genes in aerial vegetative organs of rice. This cell wall composition atlas provides a baseline for evaluating the variation of cell wall composition across developmental time points and organs, and for comparisons to other plant species. Through analysis of the correlations within and among the cell wall and gene expression datasets, we establish testable hypotheses for the potential interactions among cell wall polymers, and the functions of putative cell wall synthesis genes. For example, correlations among GAX, AGP and pectin

components suggest that the covalent interaction found in Arabidopsis may also exist in rice. We also hypothesize that AT7 and AT9 transfer FA to GAX. These results may lead to the discovery of novel grass cell wall synthesis enzymes, and may improve biofuel production and other agricultural uses of cereals and grasses.

**Results**.

This study focuses on rice aerial, vegetative organs as this grass material is used for animal feed and forage and as a biorefinery feedstock. Cell wall functions in these organs include mechanical support, transport of water and nutrients, and defense against foliar and stem pathogens and pests. We collected 30 samples of different rice aerial, vegetative organs and organ segments in 10 developmental time points of *Kitaake* rice (Figure 2-1, Table 2-S1). The 30 samples include 28 vegetative organ samples, including leaf blades, leaf sheaths, and stems; and 2 reproductive organ samples including the endosperm and hull of mature seeds. Sampling included leaves or leaf parts that represent developmental progressions from young to old, i.e., gradients (Figure 2-1 A-E and H).

We applied three complementary approaches for measuring cell wall content. First, we utilized a variety of chemical assays to determine enzymatic digestibility and 17 cell wall components, including monosaccharide residues, acetyl bromide soluble lignin, and MLG (referred to as MLG.w to distinguish it from determinations with the MLG antibody) (Figure 2-2, Table 2-S2). Second, we measured the presence of 21 polysaccharide epitopes/ligands by Comprehensive Microarray Polymer Profiling (CoMPP), an antibody-based high throughput method that detects polymers released by sequential chemical extraction of cell walls (Figure 2-S1, Table 2-S3) (Moller et al. 2007). Figure 2-3 presents the aggregated CoMPP data, summed across the three sequential extractions to represent the trend of total abundance of epitopes. The signals of the same antibody in the sequential extracts are not often correlated with each other

(Table 2-S4) since the components can bind to other cell wall polymers to different degrees, and therefore different in the resistance of solvent extraction .Therefore, we use the unaggregated CoMPP-values for correlation analysis, as described further below. Third, to confirm and extend the cell wall components measured by CoMPP, we also applied glycome profiling to extracts of three leaf samples and three stem samples (Figure 2-S2, Table 2-S5). Compared with CoMPP, the glycome profiling analysis equalizes total input sugars in each extraction and used a larger set of antibodies.

In parallel to samples for cell wall compositional analysis, we harvested samples for gene expression analysis, targeting 67 putative cell wall synthesis genes and 6 negative control genes with quantitative reverse transcriptase PCR (qRT-PCR; Table 2-S6).

We present the data as follows: 1) the compositional relationships among the samples, 2) the patterns of accumulation of the major cell wall polymers and variation in digestibility, 3) the patterns of gene expression, and 4) the correlations between gene expression and cell wall components.

***Cell Wall Content Changes During Rice Development***

Analysis of variance shows that cell wall components and digestibility vary significantly across rice development and among organs (Figure 2-2, Figure 2-3, Table 2-S2). All cell wall components, except Ara and GlcA, differ significantly among sample organ groups—young leaves, leaf blades, leaf sheaths, and stems ($\alpha$=0.05, Bonferroni corrected; Table 2-S2 - Sheet D). Digestibility and many components also

differ significantly among samples within each group, with the largest number of significantly varying components in leaf blades and young leaves (Table 2-S2 - Sheet D).

Principal component analysis (PCA) and hierarchical clustering of the chemical data indicate that most samples segregate into young and old groups (Figure 2-4A, Figure 2-S3), with the "Young Tissue" subset mostly consisting of immature organs up to 20 DPG, and the "Old tissue" subset consisting of most mature or senesced adult organs (Table 2-S1). The second principal component (PC2) separates the young and old samples. The loadings indicate that primary cell wall components, such as GalA and Rha, positively contribute to PC2, while secondary wall components, such as xylose and phenolics, negatively contribute to PC2 (Figure 2-4B). Thus, the segregation of the young and old subsets is mainly caused by the compositional difference between primary and secondary cell walls (Figure 2-4), consistent with an underlying contrast between the groups similar to the juvenile to adult transition observed in other grasses (Abedon et al. 2006). Hulls (Hul), endosperm (End) and stem nodes (Nod) cluster poorly with the young and old subsets and are therefore labeled "specialized tissue" in Table 2-S1. The cell wall composition of the two structures of "stems," nodes and internodes, differ strikingly from each other in all assays, with internodes possessing lower amounts of GAX side units and pectin components but more MLG relative to nodes (Figure 2-2, Figure 2-3 and, Figure 2-S2).

In addition to examining trends across development, we applied Pearson's and Spearman's correlation methods to the total, young, and old datasets, for the chemical assay and the unaggregated CoMPP data (Table 2-S7). Figure 2-S4 is a graphical representation of the most significant correlations for each dataset. We identified 121 significant Pearson's and Spearman's correlations in the total dataset, 61 correlations in the young subset, and 105 correlations in the old subset (Figure 2-S5, Table 2-S8). By both correlation methods, we observed a bias towards positive correlations (Figure 2-S5B), consistent with cell wall components functioning together and being connected through hydrogen bonds, covalent bonds, or simply co-regulation. Figure 2-5 represents a simplified presentation of the stronger relationships among components as indicated by a total of five or more positive correlations in any of the dataset.

***Glucuronoarabinoxylans***

The major hemicellulose in grasses, GAX components were measured by trifluoroacetate (TFA)-release of monomeric sugars (Xyl, Ara, GlcA); alkali-hydrolysis of hydroxycinnamoyl esters (FA and pCA); and the LM10 and LM11 antibodies, which recognizes unsubstituted xylans and unsubstituted and Ara-substituted xylans, respectively. Along the seedling and mature leaf blade developmental gradients, Xyl, Ara and LM10 epitopes all decrease (Figure 2-2, 2-3, and Figure 2-S2). We observe a similar trend for most of the other chemical components, which is consistent with the decrease of cell wall mass extracted (Figure 2-S6), and may be due to accumulation of silicon (de Melo et al. 2010), not measured here due to the requirement for large amounts of material (Guntzer et al. 2010). When expressed as mol%, Xyl increases as a fraction of sugars during seedling development; whereas, the Ara:Xyl ratio decreases

(Table 2-S2 – Sheet E). FA also decreases during early seedling development from 3 DPG to 9 DPG, which coincides with an increase in growth rate (Figure 2-S7). Unusual xylan substitution ratios distinguish samples classified as "specialized tissue." For example, the leaf joint (Ljn_100), which includes the leaf collar, the ligule, and the auricle (Figure 2-1N) has the highest FA, the second highest Ara content, and relatively high FA:Ara and Ara:Xyl among all samples (Figure 2-S8) , though is similar to mature leaf blade and sheath samples in many components (Figure 2-2, 2-3, 2-4).

The data exhibit many positive correlations among xylan components and with other cell wall polymers (Figure 2-5, Figure 2-S4). In the young dataset, Ara correlates with Xyl and GlcA (Figure 2-5), but in the old dataset, Xyl did not correlate with Ara but correlates with L10-xylan and *p*CA. The correlations between Ara and Xyl are dramatically different in the young and old datasets, indicative of changes in Ara substitution rates across development (Figure 2-S9). GAX components also correlate with other cell wall components such as MLG, pectins and AGPs as described below.

***Mixed Linkage Glucan***

MLG is the second most abundant hemicellulose in rice, providing tensile strength to both mature and immature tissues (Buckeridge et al. 2004; Vega-Sanchez et al. 2012). We found similar, but not identical results, from the three measures of MLG content employed, a direct lichenase assay (MLG.w), CoMPP (B-MLG), and TFA-soluble glucose (Glc), the latter of which also includes amorphous cellulose, and a small contribution from xyloglucan. Across seedling- and mature leaf-development, MLG decreases via all three measures (Figure 2-2, Figure 2-3, Table 2-S2). In mature leaf

samples, this trend is also supported by glycome profiling of the carbonate extract (Figure 2-S2). MLG measures correlate with several GAX components. For example the B-MLG epitope correlates with L11-xylan epitope in the total dataset (Figure 2-5). In addition to correlating with GAX components, the MLG epitope correlates with the CBM4-1 amorphous cellulose epitope in the total and old datasets (Figure 2-5).

***Pectins***

Pectins are the least abundant polysaccharides in grass cell walls, though their retention across evolution is consistent with important functions in grasses, especially in undifferentiated tissues (Rancour et al. 2012). The major component of pectin, GalA represents no more than 1.5% of the rice cell wall (Table 2-S2). In addition to GalA and Rha, we measured pectin with various antibodies via CoMPP (Figure 2-3) and glycome profiling (Figure 2-S2). The different pectin epitopes display different organ- and time point-enrichment patterns, suggesting that pectin functions vary by organ (Figure 2-3). L5-galactan, in particular, shows a very different binding pattern from the other pectin epitopes. Across mature leaf development (Lfb_b34, Lfb_m34, Lfb_t34), the increase in the low methylation J5-HG epitope and concomitant decrease of the high methylation J7-HG epitope suggest de-methylation of homogalaturonan during leaf development.

***Arabinogalactan Proteins***

Arabinogalactan proteins (AGPs) are glycoproteins with demonstrated roles in cell division, cell expansion, embryogenesis, root vascular tissue development and biotic stress in Arabidopsis (Majewska-Sawka and Nothnagel 2000; Nguema-Ona et al. 2014), though again there is a paucity of information for grasses. We measured AGPs with

four antibodies that recognize the oligosaccharide chains of AGP (Figure 2-3), and attribute AGPs as the origin of the monomeric sugar, Gal, though this is also a side chain component of the pectin, rhamnogalacturan I. Like those of the pectin epitopes, the AGP epitope patterns vary, but show uniformly high signals in seedlings. The data contain many correlations between components and epitopes of AGP with GAX, pectins, cellulose, and MLG (Figure 2-5, Figure 2-S4). Correlations among APG, GAX, and pectin are especially abundant in the total dataset but also present in the young data.

***Cellulose***

We estimated cellulose, the major polysaccharide in the cell wall, with the anthrone assay and two carbohydrate binding modules, CBM4-1 and CBM3a, which recognize amorphous cellulose and crystalline cellulose, respectively. These three cellulose assays measure different cellulose pools as they do not correlate with each other, except cellulose-3a and cellulose-4-1 in the old dataset. Consistent with its major structural role and fraction in the cell wall, cellulose also correlates with many other cell wall components including MLG, AGPs, and pectins (Figure 2-S6).

***Lignin and pCA***

Formed in many secondary cell walls after growth cessation, lignin is a complex phenolic polymer that in grasses includes modification with phenolic esters. Here, total phenolics represent the acetyl bromide solubilized (ABS) phenolics from total AIR, which includes feruloyl- and *p*-coumaroyl-esters (Table 2-S2). After removing these esters via alkali treatment, we also measured ABS-lignin, which represents 20 to 100%

of total phenolics (Table 2-S2 – Sheet F). Total phenolics and ABS-lignin are most abundant in hull (Hul100) (Figure 2-2), which likely contributes to physical and chemical protection of the seed against pathogens and insects (Cho et al. 1998). Mature leaf blades and stems generally have more total phenolics, lignin, and *p*CA compared to seedlings (bottom vs. top of Figure 2-2). Many correlations among lignin, total phenolics, hydoxycinnamic acids and GAX components appear in the total and old datasets (Figure 2-5, Table 2-S8). For example, total phenolics and ABS-lignin positively correlate with LM10 xylan in the total dataset.

***Enzymatic Digestibility***

A critical indicator of efficiency of biochemical conversion of biomass to biofuels, cell wall enzymatic digestibility (ED) indicates the fraction of biomass polysaccharides available for fermentation to biofuels and is highly dependent on cell wall composition (Li et al. 2015). In all datasets, the two time points (8 and 145 hours, designated ED8 and ED145, respectively) are highly correlated (Figure 2-5). The kinetics for most samples are similar, with half-times near 8 hours, though a few reactions were near completion at the early time point (Figure 2-S10). The different rice samples exhibit 7-fold variation between the highest (End100, 390 ± 90 μg/mg) and lowest (Hull100, 56 ± 15 μg/mg) enzymatic sugar release at 145 h (Figure 2-2, Figure 2-S10). The young samples are more digestible than the old samples, as is apparent with the decrease in digestibility along the seedling and leaf developmental gradients (Figure 2-2) and in the positive PC2 loadings of ED8 and ED145 in the PCA (Figure 2-4B).

Many cell wall polymers significantly correlate with digestibility. MLG.w and/or TFA-soluble Glc positively correlate with cell wall digestibility in all datasets (Figure 2-5). The L6-Arabinan epitope negatively correlate with ED in the young dataset. Lignin negatively correlates with digestibility, though weakly ($PCC_{total}$= -0.38, Table 2-S7). In agreement with the literature on biodiversity and cell wall recalcitrance (Schmer et al. 2012), ED positively contributes to PC2 along with MLG, GalA, GlcA, and Rha as opposed to the negative contribution to PC2 of Xyl and phenylpropanoids (FA, *p*CA, phenolics, and lignin, Figure 2-4B).

***Identification of Genes Involved in Cell Wall Synthesis by Correlation Analysis***

To provide insight into grass cell wall synthesis, we used qRT-PCR to measure expression of 73 rice genes, many of which were previously identified as being grass-diverged and/or highly expressed in grasses relative to dicots (Bartley et al. 2013; Cao et al. 2008; Mitchell et al. 2007). Listed in Table 2-S6, our transcript targets include those of 50 glycosyltransferase (GT), including six GT1 family members that, though highly expressed, most likely function in small molecule synthesis (Bowles et al. 2005; Lairson et al. 2008), 15 Acyl-CoA acyltransferases genes from two subclades, and 8 genes known or implicated in the monolignol biosynthesis pathway (Penning et al. 2009). To address the challenge posed by the highly interconnected and correlated nature of the cell walls for distinguishing direct and indirect correlations among analytes, we again divided samples into developmental subsets (Table 2-S9). This approach facilitated the identification of high-confidence gene-polymer correlations for nine functionally uncharacterized genes (Table 2-S10 – Sheet F).

***Gene Expression Clustering and Correlation Analysis***

We used hierarchical clustering to summarize the patterns of gene expression (Figure 2-6). Most genes grouped into five major clusters, A through E, with many genes showing high normalized relative expression (ΔΔCq) in only a few samples. Cluster A is enriched in young leaf samples like bLf13 and Lfb20 and dominated by *GT1* and *GT4* transcripts, the products of which are not directly involved in cell wall synthesis. Cluster B is highly expressed in the endosperm sample (End100) and mainly consists of *GT2, GT17* and subclade-II *AT* transcripts. Cluster C transcripts are highly expressed in Nod34 and Int_b55, where secondary wall deposition is active, but also in the leaf sample, uLf13. This cluster contains phenylpropanoid, *GT8, GT43, GT47, GT61, GT75* and subclade-I *AT* transcripts, many implicated in lignin (Gui et al. 2011; Kawasaki et al. 2006; Withers et al. 2012; Zhang et al. 2006) and GAX synthesis (Bartley et al. 2013; Chiniquy et al. 2012; Konishi et al. 2007). Cluster D transcripts are highly expressed in young samples, especially Lf3 and uLf13. This cluster consists of *AT* subclade II transcripts, GT2 transcripts including *GT2-CslfF6*, which is critical for MLG synthesis (Vega-Sanchez et al. 2012), and other *GT*s. Cluster E transcripts are highly expressed in Lsh55, Int_b55 and Lf7 and include the *GT75-UAM1* gene that indirectly affects GAX synthesis (Konishi et al. 2007). In particular, Cluster B and D provide some evidence of organ-specific expression modules, in which a set of genes appear to have sub-functionalized to act in a particular organ.

We hypothesized that the abundance of cell wall components positively correlates with expression of the corresponding synthesis gene(s) and applied both Pearson's and Spearman's correlations to this model (Table 2-S9). We also used linear regression to complement and extend the correlation analysis (Table 2-S11). We tested the analysis methods using the identification rate of known cell wall genes as an indication of "high confidence" correlations (Table 2-S12 and Table 2-S13). Pearson's correlation reproduced more high confidence correlations than Spearman's correlation, and is the focus of our discussion, as summarized in Figure 2-7 and Table 2-S10 - Sheet E. The observation that only four GT1-related correlations are found among the 92 positive significant correlations suggests that the method controls the false positive rate reasonably. We also tested a second model that applied a delay in cell wall component accumulation following expression of synthesis genes. The second model yielded fewer high confidence correlations (Table 2-S12 and Table 2-S13) and is not further discussed.

## Discussion

With the goal of revealing patterns of grass cell wall component abundance and narrowing candidate cell wall biosynthesis genes, this study examined cell wall composition, digestibility and the expression of grass-diverged and/or highly expressed cell wall candidate genes across development and organs of rice. We discuss the distributions of each cell wall polymer and cell wall digestibility among samples and correlations among polymers. Next, we highlight candidate cell wall synthesis genes revealed by the correlations between expression of known and putative synthesis genes and cell wall components.

### *Cell Wall Composition Relationships*

The GAX-related changes and correlations (Figure 2-5) during development are consistent with GAX components contributing to growth cessation and mechanical support. The increase of xylose in mol% and decrease of Ara:Xyl ratio during aging of rice seedlings is similar to seedlings in the Pooideae grass tribe, i.e., wheat, Brachypodium, and barley (Christensen et al. 2010; Gibeaut et al. 2005; Obel et al. 2002). The low Ara:Xyl ratio in old tissues may enable hydrogen-bond formation between GAX polymers or to cellulose microfibrils, thereby tethering these polysaccharides (Busse-Wicher et al. 2016; Urahara et al. 2004). Among cell wall components attributed to GAX, GlcA exhibits the strongest positive correlation with rice seedling growth rate ($PCC_{GlcA,mol\%}$ = 0.93), suggesting a model in which GlcA substitution interferes with xylan interactions with other polymers (Rennie and Scheller 2014). Examining rice mutants in orthologs of xylan GlcA-transferases would test this

(Mortimer et al. 2010; Oikawa et al. 2010; Rennie et al. 2012). The crosslinking phenylpropanoid FA, which is primarily esterified to arabinoxylan, is 10- to 50-fold more abundant and exhibits a different timing of accumulation in rice seedlings compared to Pooideae grass tribe members (Christensen et al. 2010; Obel et al. 2002). These differences may relate to disparate timing of growth between these grasses. An increase in growth rate from 5 to 9 DPG accompanies the decrease in FA during rice seedling development. Whereas, the period of rapid growth of Pooideae seedlings is earlier, when these seedlings possess low FA (Christensen et al. 2010; Gibeaut et al. 2005). The specialized cell wall of leaf joints posses high Ara and FA content, which may provide mechanical support that holds leaf blades at a characteristic angle relative to the stem (Ning et al. 2011; Xu et al. 2012).

MLG, which also plays a role in providing tensile strength, correlates with GAX components and an amorphous cellulose epitope. Interactions among MLG, GAX and cellulose have been observed in studies of primary cell walls. For example, correlation of MLG with low molecular weight GAX fragments and MLG accumulation during maize root elongation may suggest that MLG coats a portion of GAX and serves as a spacer to separate cellulose microfibrils, rendering cellular elongation irreversible (Kozlova et al. 2014). Furthermore, in primary walls absence of MLG due to mutation alters cellulose orientation and arabinoxylan-cellulose interactions (Smith-Moritz et al. 2015). Here, correlations between GAX and MLG epitopes are also apparent in the

secondary wall-rich, old dataset. Their interaction may have a role in mechanical support there as well.

AGPs in Arabidopsis have recently been found to covalently link RGI and arabinoxylan (Tan et al. 2013). The multiple correlations among AGPs, pectins and GAX suggest the possibility of similarly crosslinked structures in grasses. This is also supported by the fact that GAX fragments from grasses contain AGP or pectin sugars (Carpita 1989; Kato and Nevins 1992; Nishitani and Nevins 1989). Of course, due to the scarcity of AGPs and pectins and the high abundance of GAX, GAX-linked structures may differ between grass cell walls and those of dicots. Furthermore, it is tempting to speculate that many significant correlations between AGPs and glucans may also reflect structures yet to be revealed by careful cell wall fractionation experiments.

Correlations between phenolics and GAX components agree with the covalent bonding between GAX and lignin via ferulate esters (Bunzel et al. 2004; Ralph et al. 1995) and with the dominance of these polymers in secondary cell walls. The correlations between *p*CA and phenolics agree with the high abundance (up to 18% in maize lignin) of lignin-associated *p*CA esters (Bartley et al. 2013; Grabber et al. 2004; Hatfield et al. 2008; Molinari et al. 2013; Ralph 2010). Nonetheless, the observation that *p*CA varies significantly among organs and developmental whereas ABS-lignin does not (Table 2-S2 – Sheet F), may assist with determining the function of lignin acylation by *p*CA (Petrik et al. 2014; Ralph 2010).

Cell wall enzymatic digestibility exhibits large variation among the samples from different rice tissues and time points. Young samples are more digestible than the old

samples, consistent with efforts to "juvenilize" bioenergy crops to increase biofuel yields (e.g., (Chuck et al. 2011)). Leaves are surprisingly more recalcitrant than internodes at the same age, in contrast to the digestibility of *Miscanthus* and wheat via a *Clostridium*-based assay (Costa et al. 2014). This may be due to the relatively high silicon content of rice leaves (Yamaji et al. 2008), and reflected by the low measured cell wall content in leaves versus stems (Figure 2-S6).

Cell wall digestibility positively correlates with MLG and negatively correlates with an arabinan epitope and lignin. The correlation between digestibility and MLG is consistent with the observation of a weak positive correlation between MLG and ED among diverse rice genotypes grown in both the greenhouse and field (Tanger et al. 2015). This relationship may be due to MLG loosening the structure of cell walls and increasing the accessibility of cellulose (Kozlova et al. 2014), but also our observation that MLG, itself, is digested by the cellulase and beta-glucosidase cocktail used. In either case, this correlation supports the notion that MLG is a preferred cell wall component for a biofuel feedstock (Tanger et al. 2015). The negative correlations of the LM6-Arabinan epitope and ED in the young dataset are consistent with a role for pectinases in improving switchgrass cell wall digestibility (Chung et al. 2014). Lignin negatively correlates with digestibility, though weakly. The weaker than expected correlation may be due to the vastly different cell wall structures of the tissues examined, and is in agreement with emerging evidence that lignin is not as important a determinant of recalcitrance in grass biomass as it is in tree biomass (De Souza et al. 2015; DeMartini et al. 2013).

***Identification of Genes Involved in Cell Wall Synthesis by Correlation Analysis***

We measured expression of selected grass-diverged and highly expressed *GT*s and *AT*s and observed many correlations between the expression of known/putative synthesis genes and cell wall components, which lead to identification of nine candidate cell wall synthesis genes (Table 2-S10 – Sheet F).

Four genes, two encoding GT75-UAMs and two encoding GAUT-like proteins, are "hubs" in the cell wall component-gene network, each correlating with at least five components (Figure 2-7, Table 2-S10, Sheet E). About half of the *GT75-UAM1* and *GT75-UAM3* correlations are with GAX, AGP, and pectin components. GT75-UAM1 and GT75-UAM3 are UDP-arabinopyranose mutases that convert UDP-arabinopyranose to UDP-arabinofuranose, which is used for the synthesis of GAX, the polysaccharide of AGPs, and arabinogalactan sidechains of rhamnogalacturonan I (Rautengarten et al. 2011). Since Ara is connected to many other moieties of these polymers, other components are also affected by the expression of *UAM*s. For example, down regulation of *GT75-UAM1 and GT75-UAM3* decreases both FA and Ara in rice (Konishi et al. 2011; Konishi et al. 2010). In fact, many *GT75-UAM1* and *GT75-UAM3* correlations are with GAX, AGP, and pectin components, consistent with their functions in synthesizing precursor sugars for cell wall synthesis. However, the correlations between *GT75-UAM3* and MLG.w and Glc remain to be understood. The other closely related member *UAM2*, which lacks mutase activity (Konishi et al. 2010), does not exhibit correlations above our stringent criteria, as expected.

The other two hub genes, *GT8-GAUT1L* and *GT8-GAUT9L,* significantly correlate with pectin and AGP components (Figure 2-7). The correlation between *GT8-GAUT1L* and GalA and other pectin/AGP components suggest that this rice gene may also encode an HG:GalA transferase that synthesizes the backbone of HG as the related GAUT1 does in Arabidopsis (Atmodjo et al. 2013; Caffall et al. 2009). *GT8-GAUT9L*, a putative co-ortholog of *GAUT9*, also correlates with GalA, consistent with the decrease in GalA in *gaut9-3*, an Arabidopsis GAUT9 knockdown mutant (Caffall et al. 2009), though the acceptor of GAUT9 remains to be determined. In summary, our study indicates that the function of GAUTs 1 and 9 orthologs in rice are likely to be conserved between Arabidopsis and the grasses, though work is required to test this and determine onto which oligosaccharides they transfer GalA.

Several GTs are implicated in GAX synthesis (Figure 2-7). *GT61-XAX1L-1* correlates with FA in the old subset and with the LM13-pectin epitopes in the total dataset; *GT61-XAX1L-1* also significantly relates to FA by linear regression. This suggests that, like other GT61 proteins (Anders et al. 2012; Chiniquy et al. 2012), *GT61-XAX1L-1* also participates in xylan synthesis, such as by adding side groups to GAX that are (or will be) feruloylated. *GT77-4* and *GT17-C-1* also significantly correlate with GAX components, in the total dataset, though these gene families have not been implicated in GAX synthesis. Both correlations are supported by linear regression results (Table 2-S11). The encoded proteins may function in the addition of minor GAX side groups (Rennie and Scheller 2014) or in processing of xylan synthesis enzymes (Olszewski et

al. 2010), respectively. Distantly related Arabidopsis GT77s function as xylosyl and arabinosyltransferases involved in RGII synthesis and protein glycosylation, respectively (Egelund et al. 2006 ; Gille et al. 2009). *GT10-FucTAL* and *GT2-CslA6* also correlate with GAX components but are not supported by linear regression results. They may synthesize cell wall components that correlate with the GAX components. In addition, we might have expected to observe a correlation between *OsIRX9* and GAX components. However, since several rice GT43s function in GAX backbone synthesis (Chiniquy et al. 2013; Lee et al. 2014b), a direct correlation between this individual GT43 and xylan abundance may have been obscured.

Three GT2s may be involved in MLG synthesis. Each of the different measures of MLG correlates most strongly with expression of a different gene implicated in MLG biosynthesis (Figure 2-7). GT2-CslF6 correlates with the B-MLG epitope in CDTA extracts in the total dataset; GT2-CslH1 correlates with TFA-soluble Glc in the total dataset; and GT2-CslF8 correlates with lichenase-based MLG.w in the young dataset. Based on heterologous expression, the current model of MLG synthesis is that CSLF and CSLH enzymes function independently of each other and are each capable of synthesizing MLG (Burton et al. 2006; Doblin et al. 2009; Vega-Sanchez et al. 2012). Genetic studies of rice *csfl6* found a ≥97% decrease in detectable MLG content in the *cslf6* knock out mutant (Vega-Sanchez et al., 2012); whereas, a barley *CslF6* missense mutation reduces grain MLG by 60% (Hu et al. 2014). A quantitative trait locus for MLG in barley finds that *HvCslF8*, the ortholog of rice *CslF8*, and three other *HvCslF* map to a region that causes ~19% of the variation of MLG in grain (Burton et al. 2008).

Though recent studies discount the participation of CslH1 in MLG biosynthesis in seedling organs based both on protein localization and gene expression (Wilson et al. 2015), the *CslH1* gene of rice is reasonably expressed (Cao et al. 2008). The different highly significant predominant correlations of each of these genes, and the existence of multiple gene family members in the CslF and CslH families suggest spatial and temporal regulation of MLG synthesis that merits additional genetic and biochemical studies of *CslF* and *CslH* gene family members in rice.

We also identified numerous correlations between cell wall components and so-called "Mitchell-clade" acyltransferases that modify cell wall polymers with HCAs (Bartley et al. 2013; Buanafina et al. 2016; Mitchell et al. 2007; Sibout et al. 2016) (Figure 2-7, Table 2-S10). Two studied members of this clade in rice are *p*CA transferases, with OsPMT/OsAT4 and the *Brachypodium distachyon* PMT, transferring *p*CA to monolignols (Petrik et al. 2014; Withers et al. 2012) and rice AT10 functioning in *p*CA esterification on a five-carbon sugar, likely Ara of GAX (Bartley et al. 2013). It has also been suggested that some AT subclade I homologs in *Brachypodium* may be FA-transferases involved in cell wall modification (Buanafina et al. 2016; Molinari et al. 2013). In our results, *OsPMT/OsAT4* positively correlates with *p*CA as expected (Petrik et al. 2014; Withers et al. 2012). *OsAT10* positively correlates with a xylan epitope though a correlation between *AT10* expression and total *p*CA is not observed, which may be due to the relatively low abundance of this modification. *OsAT7* positively correlates with a xylan epitope, L10-xylan-C, and may be involved in HCA incorporation onto GAXs. In a screen a rice *AT7* knock-down mutant showed decreased

FA-esters in rice leaf sheath cell walls, reinforcing this hypothesis (Bartley et al. 2013). The correlation between *OsAT9* and FA in the old dataset suggests rice AT9 may also be a FA-arabinosyl transferase. Other uncharacterized acyltransferses like AT12, AT14, AT16, AT17, AT18 correlate with other cell wall components in different datasets. Besides lignin and GAX, feruloylation of other polymers has been documented in grasses or dicots, including xyloglucan, arabinan, and galactan (Ishii and Hiroi 1990; Mathew and Abraham 2004). Based on the dispersion of ATs with significant correlations to various cell wall components (Figure 2-6), it is tempting to suggest that there may be other instances of cell wall hydroxycinnamylation in rice that remain to be revealed.

## Conclusion

The cell wall composition profile of rice presented here provides a detailed atlas for understanding the distribution of chemical components and cell wall epitopes across aerial vegetative development. We applied both chemical and antibody-based assays to quantify cell wall polymers and, in many cases, found significant correlations between chemical components and antibody/CBM-recognized cell wall epitopes. On a technical level, correlations among related components measured with different assays, such as the MLG epitope, MLG.w lichenase measurement, and TFA-soluble glucose indicate that the relative quantitation of many epitopes agrees with the abundance of components as measured by chemical assays. The correlations among cell wall components such as GAX, AGPs and pectins also suggest possible covalent or non-covalent interactions that should be tested by further chemical and structural analysis. The cell wall data will facilitate further studies of rice and other grasses, providing guidance for when and where samples should be collected in order to observe phenotypes related to particular cell wall polymers. We have also revealed potential physiological functions for grass cell wall substituents, such as the observation that GlcA correlates with high seedling growth rate.

In addition to providing insight to cell wall development and function, this map lays a foundation for developing hypotheses for the biochemical and physiological function of genes that exhibit coordinated expression with specific cell wall components. This work successfully moves beyond the observation that particular GTs and ATs are grass-diverged and highly expressed to providing testable hypotheses for specific cell wall

synthesis activities, pinpointing nine uncharacterized genes with putative functions in rice cell wall synthesis. These results indicate that the correlation analysis approach presented here is valuable for identifying novel synthesis genes that cannot be identified via simple phylogenetic analysis. This method can also be used to identify genes likely to function in cell wall and other metabolite synthesis in whole-genome transcriptome data. However, to be most effective, such an analysis will require many samples with high cell wall composition variation to identify significant correlations among tens of thousands of genes.

## Materials and Methods

Supplemental File 1 contains more detailed information.

### *Rice Growth Conditions, Measurements, and Harvest.*

The names of samples are defined by organ, segment, and age of plant in days post germination (i.e., sowing) as described in Table 2-S1. *Oryza sativa* ssp. *japonica* cv. Kitaake seedlings were grown in topsoil in a greenhouse. Harvests of healthy, developmentally similar plants with a uniform height were conducted between 2:30 and 6:00 pm with samples frozen on dry ice immediately. Six biological replicates or replicate pools were harvested for each sample, and three each used for cell wall and RNA analysis.

### *Cell Wall Preparation and Measurements*

For all cell wall assays, fresh-frozen material were ground, made into AIR and destarched as described previously (Bartley et al. 2013). Assay methods are summarized in Table 2-S2 – Sheet A. "Total Phenolics" represents the traditional acetyl bromide soluble lignin of the destarched AIR (Fukushima and Hatfield 2004), scaled for a micro-plate reader, as described (Bartley et al. 2013). "ABS lignin" was measured on the pellet remaining from the HCA extraction protocol. Chemical measures were made in biological triplicates. CoMPP was performed as described previously (Moller et al. 2007) on 15 to 50 mg dsAIR pools created from biological replicates with technical triplicates, except for the Lf3 sample, for which duplicates were performed. Antibodies or carbohydrate binding modules (CBMs) used are listed in Table 2-S3-Sheet A. Z-scores were calculated as the biological average for a given sample minus the mean of the biological average values for all the samples, divided by the standard deviation of

all the biological averages. Glycome profiling was conducted as reported previously (Pattathil et al. 2012) with 50 mg dsAIR pools of the three biological replicates per sample.

***Gene Expression Measurements by Quantitative PCR***

The method of RNA preparation, cDNA synthesis, and quantitative PCR using SYBR green was as described (Bartley et al. 2013) with some exceptions. The *Ubq5* primer pair (Jain et al. 2006)was used for normalization (ΔCq) and a pool of all cDNAs was run on each plate to determine relative expression (ΔΔCq). Samples for which the coefficient of variation of ΔΔCq among replicates was >2 were excluded from analysis. Primers used for quantitative PCR are listed in Table 2-S14.

***Data Analysis***

R 2.15.2 (http://www.r-project.org/) was used for all analyses unless otherwise stated. The "prcomp" function was used for Principle components analysis (PCA), and the "pvclust" function was used for hierarchical clustering of the cell wall chemical data. Approximately unbiased p-values for the null hypothesis that the clade is not found in any of the rejected trees were calculated from 1000 multiscale bootstrap resampling. To identify significant differences in cell wall components within and between sample groups, we used ANOVA followed by Tukey's range tests with a significance level ($\alpha = 0.05$) adjusted via a Bonferroni family-wise error rate correction. Chi-squared tests on the number of positive and negative correlations for each dataset were performed in Excel.

We analyzed Pearson's and Spearman's correlations (PCC and SCC, respectively) within the cell wall data and between the cell wall chemical components, disaggregated CoMPP data, and gene expression data for the total, young, and old datasets. Pearson's and Spearman's correlation coefficients and associated *p*-values and *q*- values were calculated with the "CCA" and "qvalue" packages (Storey and Tibshirani 2003). Because Pearson's correlation is especially sensitive to outliers, we also applied a robustness filter to remove cases in which the correlation is determined largely by a single outlier point. For correlations within the cell wall data, we used $q<0.001$ for the total dataset, $q<0.02$ for the young and old datasets, and a $|\Delta CC|_{Max}<0.5$ for all datasets. For correlations between cell wall components and genes with the first model (N vs. N) we used $q<0.07$ and $|\Delta CC|_{Max}<0.4$ for all datasets, along with the following criteria: $|CC|>0.66$ for the total, $|CC|>0.90$ for the young, and $|CC|>0.84$ for the old dataset. Different Correlation coefficient cutoffs are applied due to the varying numbers of samples in each data subset.

For linear regression analysis between gene expression data and cell wall data, significant linear regressions were identified using a *q*-value <0.07 for all datasets.

## Figures and Tables

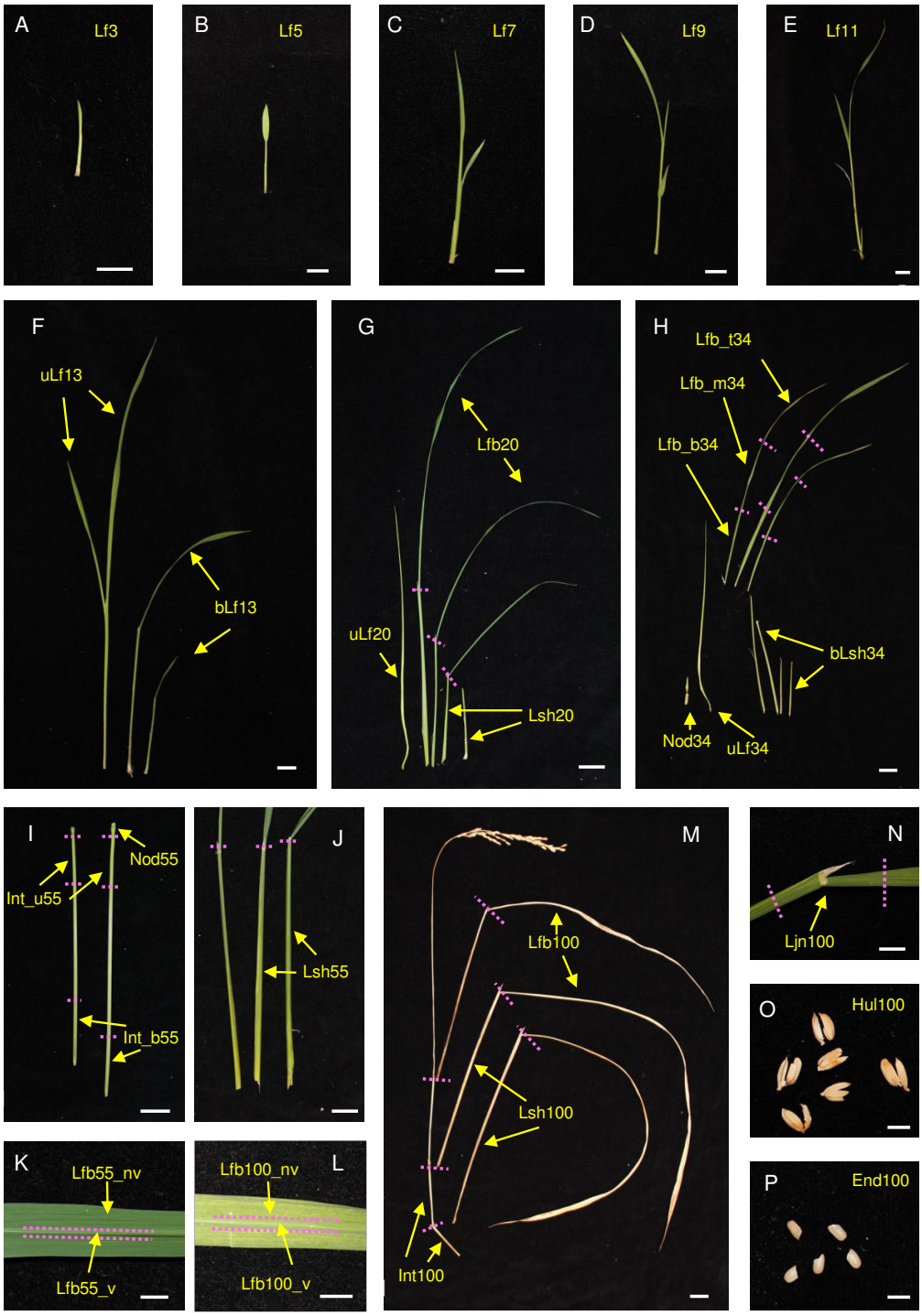



**Figure 2-1. Rice samples characterized in this study.** (A-E) Rice seedlings from 3 day-post-germination (DPG) to 11 DPG. (F) Dissected plant at 13 DPG. (G) Dissected

plant at 20 DPG (H) Dissected plant at 34 DPG, Lfb_b34, Lfb_m34, and Lfb_t34 are samples representing mature leaf developmental gradient from young to old. (I) Dissected stems at 55 DPG. (J) Dissected leaves at 55 DPG. (K) Dissected leaf blades at 55 DPG. (L) Dissected leaf blades at 10 DPG. (M) Dissected plants at 100 DPG. (N) Leaf joint at 100 DPG. (O) Seed hulls at 100 DPG. (P) Seed without endosperm at 100 DPG. Scale bars indicate 1 cm in (A-F), indicate 2 cm in (G-J), (M) and indicate 0.5 cm in (K), (L), (N-P). The abbreviations of samples are labeled in yellow. Magenta lines indicate positions of dissection. More detailed description of samples can be found in Table 2-S1.

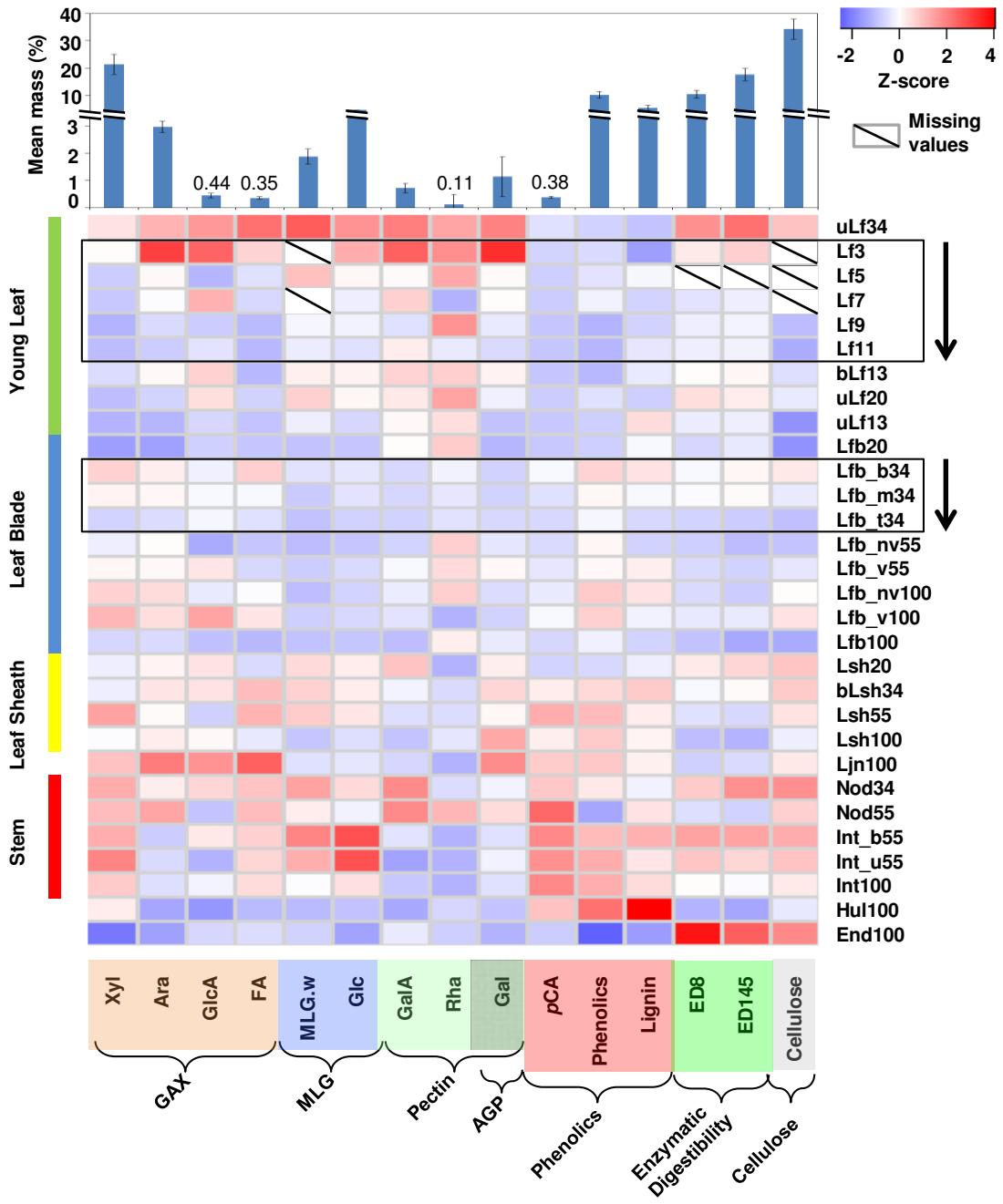



**Figure 2-2. Abundance of cell wall components and properties measured by chemical assays.** Color intensity indicates the z-score of abundance for each component. Red indicates relative enrichment and blue, relative scarcity. Slashed squares are missing values due to insufficient sample mass. The mean, un-normalized

mass for each component is shown above the heatmap. The color ribbons on the left of heatmap indicate organs. Samples are arranged by organs and in an ascending order of maturity (i.e., young to old). Black boxes indicate developmental series with the arrows on right indicating the direction of the order of maturity. The color ribbons at the bottom demarcate categories of cell wall components. Sample and components abbreviations are described in Table 2-S1 and Table 2-S2.

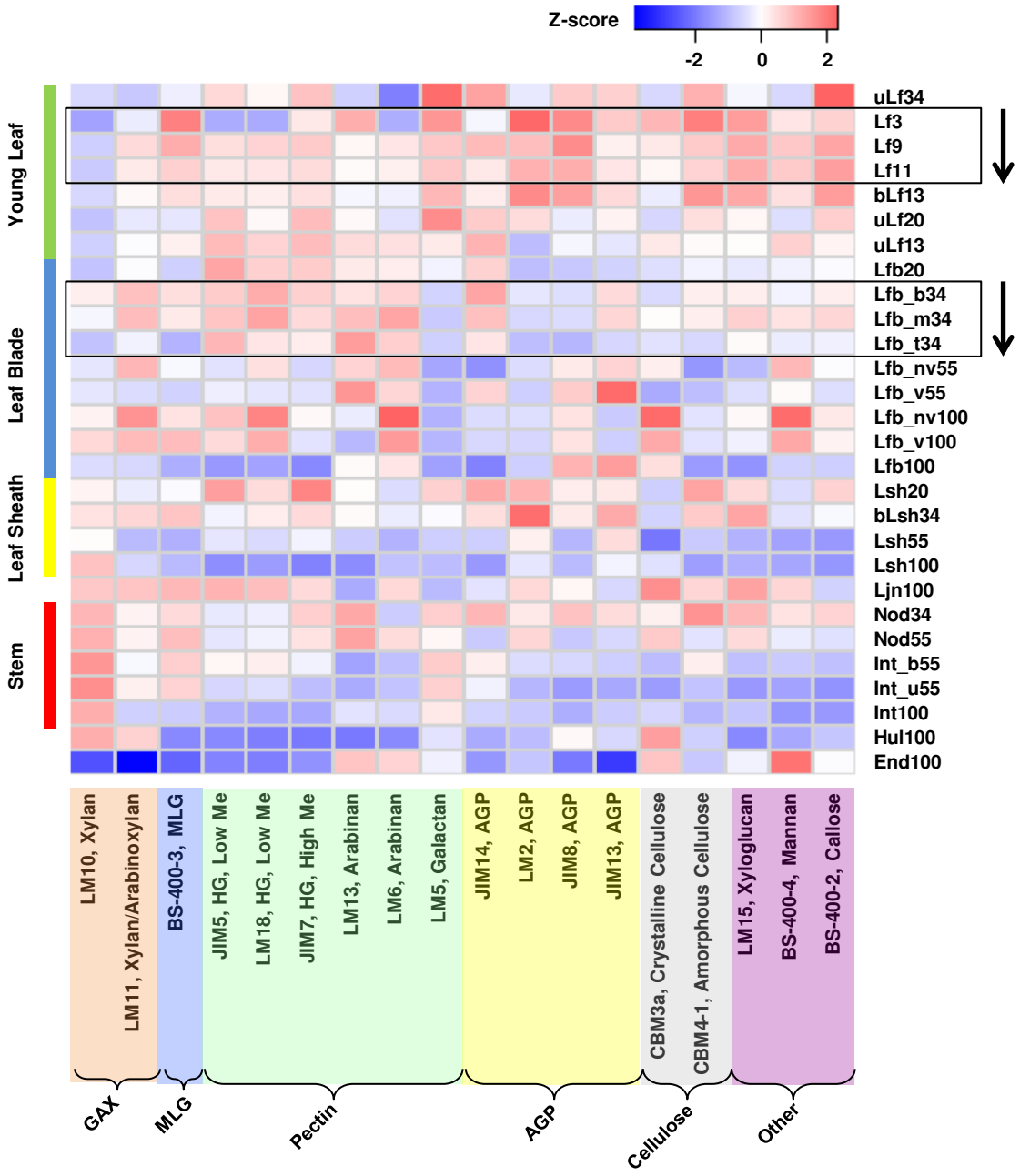



**Figure 2-3. Comprehensive Microarray Polymer Profiling (CoMPP) analysis of cell wall epitopes and ligands.** Color intensity indicates the z-score of abundance of cell wall epitopes summed across the three extractions. Red indicates relative

enrichment and blue, relative scarcity. Samples are arranged by organs and in an ascending order of maturity. The color ribbons on the left of heatmap indicate organs and those at bottom indicate categories of cell wall components. Sample and epitope and ligand abbreviations are described in Table 2-S1 and Table 2-S3.

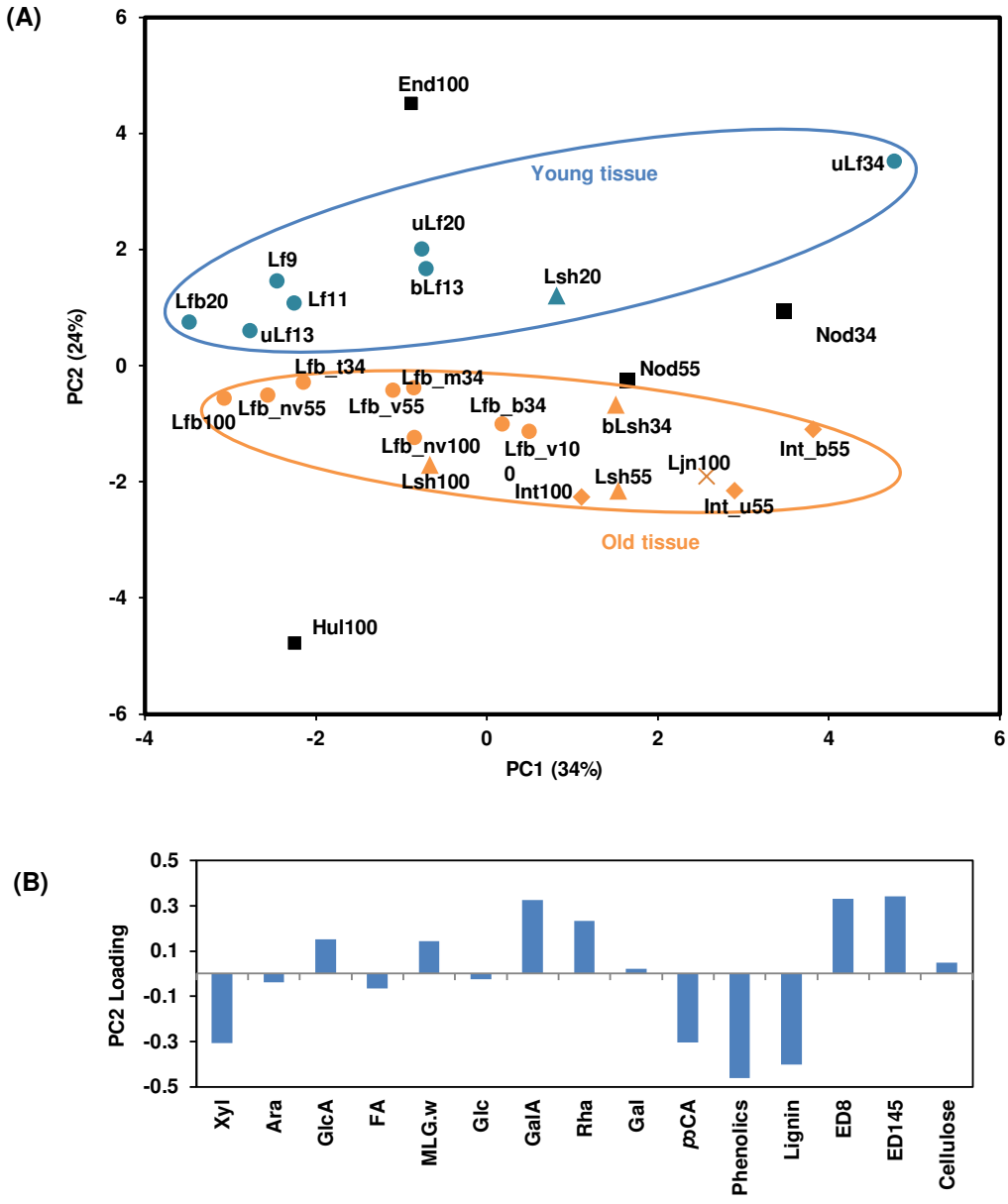



**Figure 2-4. Principal component analysis of chemical assay dataset shows two sample groups by Principal Component 2.** (A) Axes values indicate principal components score and the percentages indicate the variation represented by the

principal component. Green symbols indicate young tissue, orange symbols indicate old tissue, and black symbols indicate samples that are not easily classified. Symbol shape indicates the organ: circle for leaf blade, triangle for leaf sheath, diamond for internode, X for leaf joint, and square for specialized organs. Sample abbreviations are described in Table 2-S1. (B) The loadings of PC2. Cell wall component abbreviations are described in Table 2-S2.

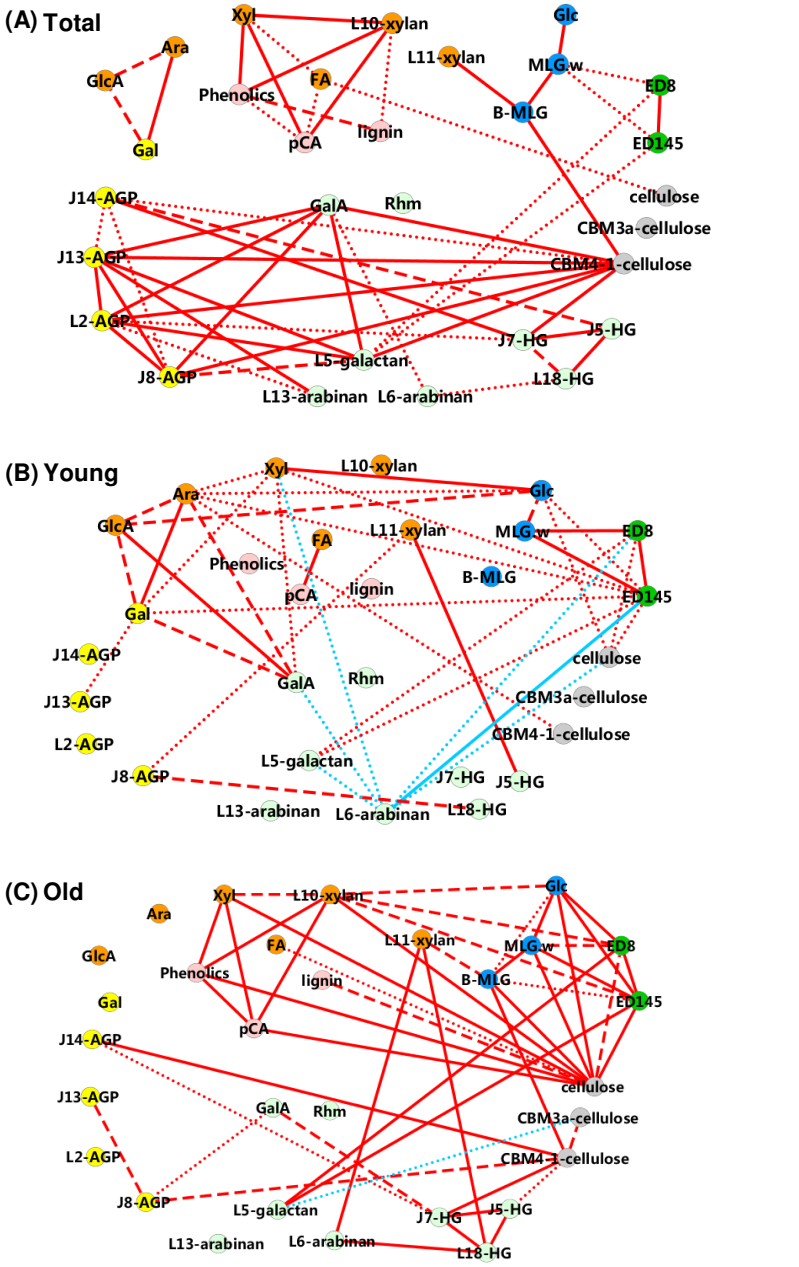

(A) Total
Xyl
L10-xylan
Glc
Ara
L11-xylan
GlcA
MLG.w
ED8
Phenolics
FA
B-MLG
lignin
pCA
ED145
Gal
J14-AGP
GalA
Rhm
cellulose
CBM3a-cellulose
J13-AGP
CBM4-1-cellulose
L2-AGP
J7-HG
J5-HG
L5-galactan
J8-AGP
L18-HG
L13-arabinan
L6-arabinan
(B) Young
Xyl
L10-xylan
Ara
Glc
GlcA
FA
L11-xylan
MLG.w
ED8
Phenolics
B-MLG
pCA
lignin
ED145
Gal
J14-AGP
cellulose
GalA
Rhm
J13-AGP
CBM3a-cellulose
CBM4-1-cellulose
L2-AGP
L5-galactan
J8-AGP
J7-HG
J5-HG
L13-arabinan
L6-arabinan
L18-HG
(C) Old
Xyl
L10-xylan
Glc
Ara
FA
L11-xylan
MLG.w
ED8
GlcA
Phenolics
lignin
B-MLG
ED145
Gal
pCA
J14-AGP
cellulose
J13-AGP
GalA
Rhm
CBM3a-cellulose
L2-AGP
CBM4-1-cellulose
J8-AGP
L5-galactan
J7-HG
J5-HG
L13-arabinan
L6-arabinan
L18-HG
Key:
Both PCC and SCC
Only PCC
Only SCC
Positive Correlation
Negative Correlation
Cellulose
GAX
Pectin
Lignin
MLG
AGP
Enzymatic Digestibility

**Figure 2-5. A simplified representation of the significant correlations among cell wall components**. Correlations with (A) total, (B) young, and (C) old datasets. Complete correlation networks are shown in the Figure 2-S4. To simplify the network ，we only demonstrate the correlations discussed and combine the three extractions of CoMPP to one node and an edge was added if any of the three extraction values correlate with another component. Xyloglucan, mannan, and cellulose correlations were excluded. Red and cyan lines indicate positive and negative correlations, respectively. The line style indicates whether the relationship is significant via Person's correlation coefficient (PCC) and/or Spearman's correlation coefficient (SCC) as indicated by the key. The thickness of lines is proportional to the absolute value of the higher PCC or SCC correlation. Cell wall components are represented by circular nodes, color-coded based on the cell wall polymers from which they originate, as described in Table 2-S3. Abbreviations used are defined in Table 2-S2 and Table 2-S3.

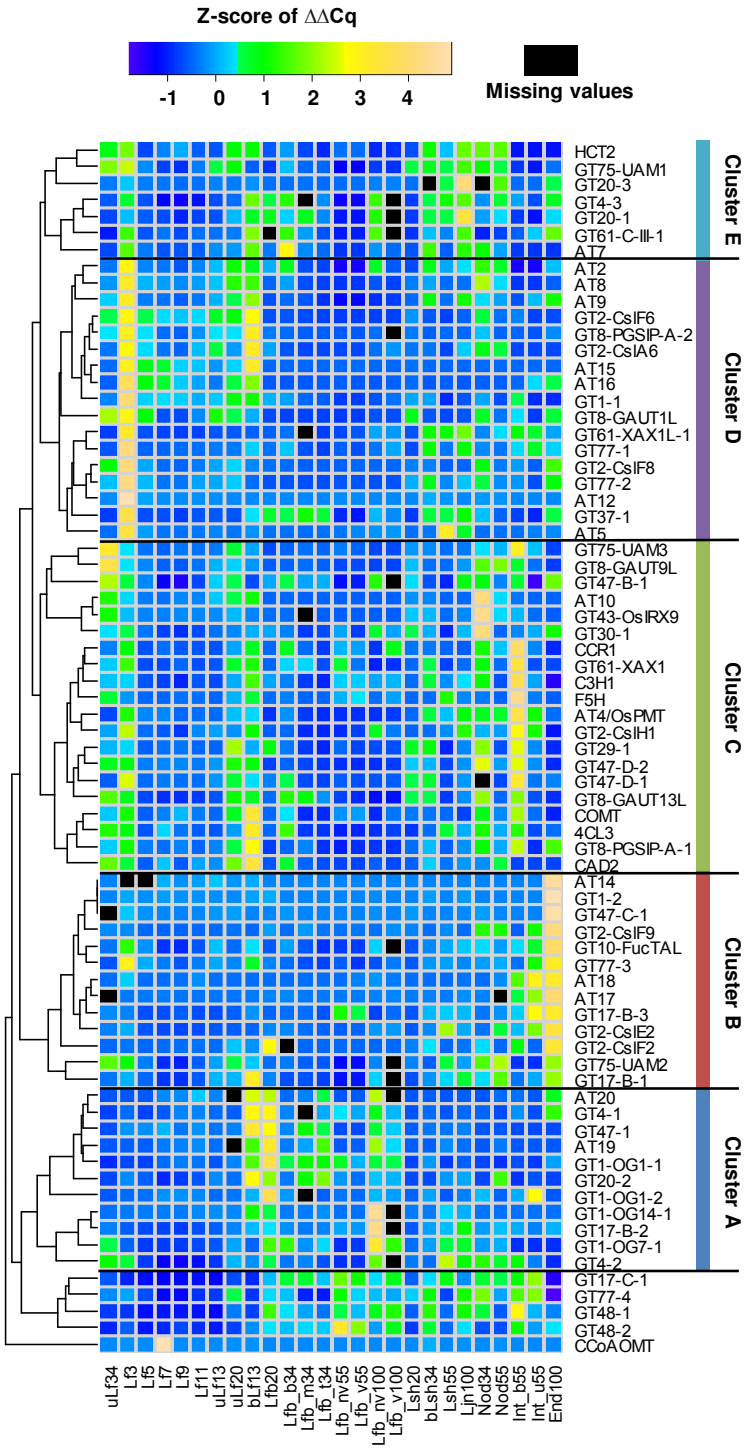



**Figure 2-6. Gene expression profiles measured via qRT-PCR.** The normalized relative expression of each gene is represented by the Z-score of ΔΔCq. The major gene

clusters were determined by hierarchical clustering (N=5000) shown as a dendrogram on the left. Gene name codes are described in Table 2-S6.

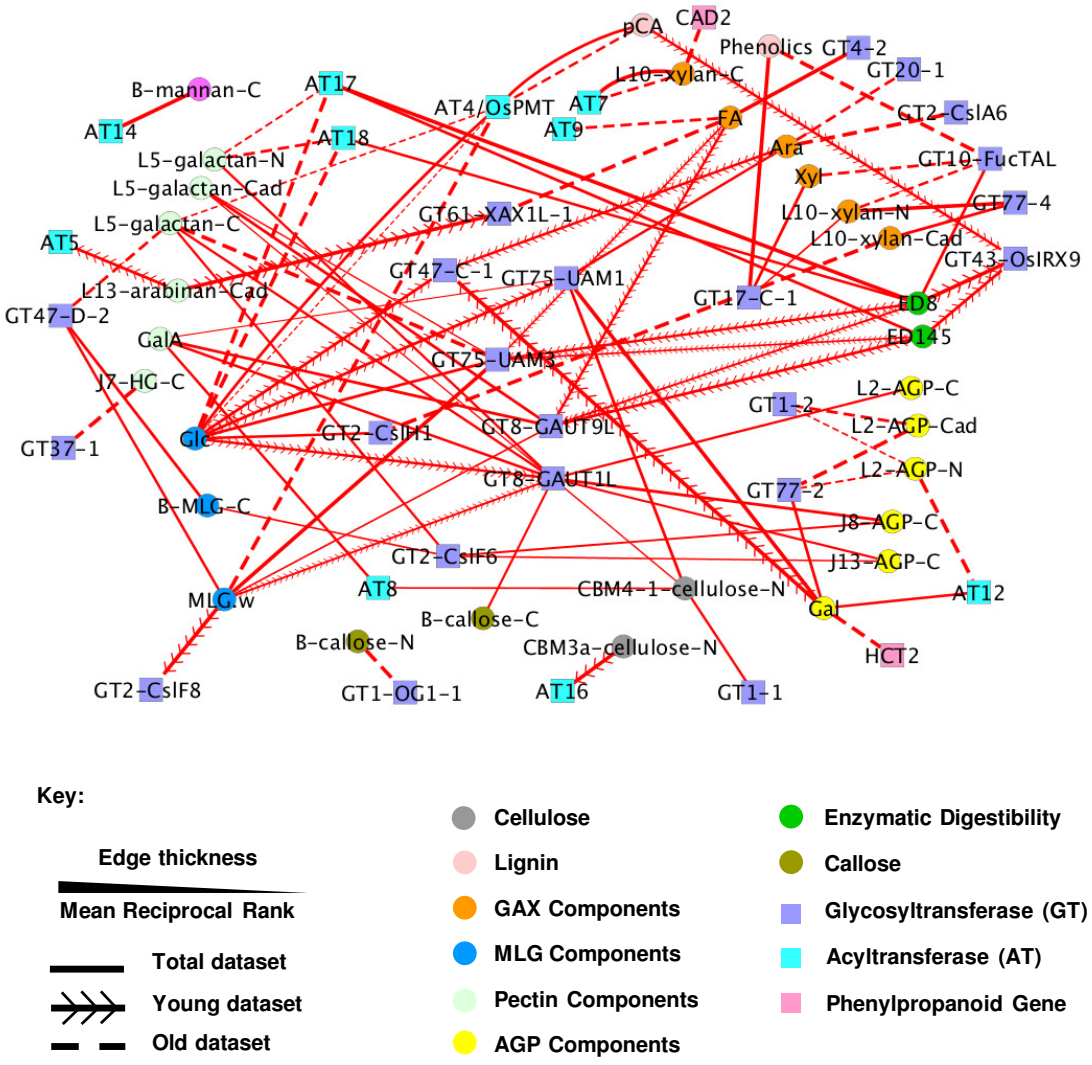



**Figure 2-7. Network of genes that positively correlate with cell wall components.** Significant Pearson's correlations were identified by setting cutoffs of *q*-value, correlation coefficient, and robustness of correlations as stated in the method. The styles of line indicate the correlations in different datasets as indicated by the key. The thickness of lines is proportional to the rank of components among all significant

correlated components. Cell wall components shown by oval nodes are color coated based on the cell wall polymers they mainly come from as shown in Table 2-S3. Squares indicate genes and are color-coded based on their protein family. Gene names are provided in Table 2-S6. Cell wall components and epitope names are provided in to Table 2-S2 and Table 2-S3.

***Supplementary data***

Additional supplementary data are available at PCP online.

https://academic.oup.com/pcp/article/57/10/2058/2755865/Cell-Wall-Composition-and-Candidate-Biosynthesis#50717385

**Supplementary file 1:** Supplemental Methods.

**Supplementary file 2:** Comprises ten additional figures that complement the data presented in the manuscript. **Figure 2-S1.** Relative abundance of cell wall epitopes and ligands measured by CoMPP. **Figure 2-S2.** Relative abundance of cell wall epitopes in rice stem and old leaf samples by glycome profiling. **Figure 2-S3.** Hierarchical clustering of samples based on cell wall composition. **Figure 2-S4.** Significant correlations among cell wall components in the (A) total, (B) young, (C) old datasets. **Figure 2-S5.** Venn diagrams of the classes of significant correlations among cell wall components. **Figure 2-S6.** The total fraction of cell wall materials measured in alcohol insoluble residue of each sample. **Figure 2-S7.** Plant height and growth rate for early rice seedlings. **Figure 2-S8.** Ratios of glucuronoarabinoxylan components across samples. **Figure 2-S9.** The relationship between arabinose and xylose differs among data subsets. **Figure 2-S10.** Enzymatic digestibility across all samples.

**Table 2-S1:** Nomenclature, description, and categories of rice samples characterized in this study. Samples are referred to by abbreviations for the organs; the segment, if applicable; and the age of plant in days post germination (DPG).

**Table 2-S2:** Description of cell wall chemical composition assays and digestibility, corresponding data, and ANOVA analysis for samples and sample groups.

**Table 2-S3:** Antibodies and carbohydrate-binding modules (CBMs) used to detect cell wall epitopes with CoMPP and raw CoMPP data.

**Table 2-S4:** Correlations among extracts of a given epitope in CoMPP in different data subsets.

**Table 2-S5:** Glycome profiling optical density values and sugar recovered from each extract.

**Table 2-S6:** The categories and annotations of genes probed in this study and their normalized relative expression (ΔΔCq) determined by qPCR.

**Table 2-S7:** Correlation coefficients and q-values among all cell wall components.

**Table 2-S8:** Significant correlations among cell wall components.

**Table 2-S9:** Pearson and Spearman correlation coefficients and q-values between cell wall variables and transcripts.

**Table 2-S10:** Significant positive correlations between cell wall components and transcripts in the N vs. N model.

**Table 2-S11:** Linear regression associations between gene expression and cell wall components.

**Table 2-S12:** Significant positive correlations between cell wall component and transcript amount with the N vs. Δ(N+1) model and the sample pairs used.

**Table 2-S13:** Comparison of Pearson's correlation (PCC) and Spearman's correlation (SCC) for model 1 (N vs. N) and model 2 (N vs Δ(N+1)).

**Table 2-S14:** Primers used for qRT-PCR.

# Chapter 3: Proteome, Metabolite, and Cell Wall Profiling of an Elongating Rice Stem Internode, a Secondary Cell Wall Biosynthesis Model

**Authors:** Fan Lin, Brad J. Williams, Athena A. Schepmoes, Stephen J. Callister, Padmavathi A. V. Thangella, Adam Ladak, Hernando J. Olivos, Kangmei Zhao, Laura E. Bartley

**Publication Status**: This chapter has been submitted to Frontiers in Plant Science.

**Author contributions:**

LB, PT, and FL conceived of this study and designed the experiments. PT and FL prepared samples for proteome and metabolite profiling respectively. BW performed LC-MS/MS and data processing for experiment 2. AS and SC performed LC-MS/MS and data processing for experiment 3. AL, and HO performed LC-MS for metabolite profiling. FL and KM analyzed and interpreted the data; FL and LB drafted the manuscript. All authors revised the manuscript critically and approved the final version for publication.

**Abstract**

Internodes of grass stems function in mechanical support, transport, and, in some species, are a major sink organ for carbon in the form of cell wall polymers. To establish the rice elongating internode as a model for secondary cell wall development, we conducted cell wall composition, proteomic and metabolite analyses. We measured secondary cell wall components along eight segments of the second rice internode (internode II) at booting stage. Cellulose, lignin, and xylose increase as a percentage of cell wall material from the younger to the older internode segments, indicating active cell wall synthesis. For whole elongating internodes, we measured trypsin-digested peptides of size-fractionated proteins via liquid-chromatography tandem mass spectrometry (LC-MS/MS). This study identified a total of 2356 proteins with at least two unique peptides, including many glycosyltransferases, acyltransferases, glycosylhydrolases, cell wall-localized proteins, and protein kinases that have or may have functions in cell wall biosynthesis or remodeling. We also identified 21 unique phosphopeptides belonging to 20 phosphoproteins including an LRR-III family receptor like kinase. GO over-representation analysis and KEGG pathway analysis indicate many proteins involved in biosynthetic processes, especially the synthesis of secondary metabolites such as phenylpropanoids and flavonoids. We also used LC-MS to measure hot methanol-extracted secondary metabolites from whole internodes at the elongation stage, early mature stage, mature stage and post mature stage, and, for comparison, leaves and roots at the mature stage. The results indicate secondary metabolites in stems are distinct from those of roots and leaves, and differ during stem maturation. This study fills a void of knowledge of proteomics and metabolomics data for grass stems,

specifically for rice, and provides baseline knowledge for more detailed studies of cell wall synthesis and other biological processes during internode development. This and future work is aimed at optimizing stem development and cell wall composition of grasses to improve agronomic properties.

## Introduction

The grass family, Poaceae, includes the cereal crops and represents one of the most wide-spread plant taxonomic groups in terrestrial ecosystems (Kellogg 2001). Grass stems, also known as culms, mechanically support reproductive structures, transport nutrients and act as structural and non-structural carbohydrate storage organs (Moldenhauer et al. 2013). In rice, stem internodes rapidly elongate at the beginning of the reproductive stage (Bosch et al. 2011; Slewinski 2012). Many important biological processes including cell division, cell wall synthesis, and cell wall remodeling occur during stem development (Bosch et al. 2011; Cui et al. 2012). Among these processes, the change in cell walls is especially important for stem mechanical properties (Gritsch and Murphy 2005; Wang et al. 2012). The developmental stage of stems influences susceptibility to pests and pathogens (Bandong and Litsinger 2005; Viajante and Heinrichs 1987 ).

Though overall culm development is acropetal, with younger internodes at the top away from the roots, each grass internode exhibits basipetal development. The intercalary meristem, located at the bottom of each internode, undergoes cell division to produce new cells that only have primary cell walls. The new cells elongate and then gradually mature, forming secondary walls toward the apex of each internode (Kende et al. 1998). Associated with this developmental progression, or gradient, is a change in cell wall composition dominated by deposition of secondary cell walls, which are deposited between the plasma membrane and primary walls in some cell types such as fibers and sclerenchyma cells. As with those of dicotyledonous plants, secondary cell walls of

grasses usually contain multiple layers with highly oriented cellulose microfibers, each at different orientations. The secondary walls are also layered or impregnated with the covalently crosslinking phenylpropanoid-derived lignin polymer and have higher strength, but lower extensibility relative to primary walls. In maize, secondary cell wall components such as cellulose, lignin, and a major grass hemicellulose, arabinoxylan, increase from the bottom to the top part of each internode, while other cell wall components, such as mixed-linkage glucan, mannan and pectins decrease (Zhang et al. 2014). These changes are associated with the abundance of transcripts for cell wall synthesis enzymes, cell wall remodeling enzymes, and associated regulatory proteins (Zhang et al. 2014).

Many proteins function in cell wall synthesis and remodeling, or regulate the amounts, localization, or activity of cell wall enzymes. Members of several glycosyltransferase (GT) families, including GT2, GT43, GT47, and GT61, synthesize cellulose and hemicellulose in grasses (Scheible and Pauly 2004; Scheller and Ulvskov 2010). Enzymes in the phenylpropanoid pathway synthesize lignin precursors and hydroxycinnamic acids (HCAs). The latter are incorporated into lignin or polysaccharides by so-called BAHD acyl-CoA acyltransferases (ATs) (Bartley et al. 2013; Buanafina et al. 2016; Lin et al. 2016; Molinari et al. 2013; Petrik et al. 2014; Sibout et al. 2016; Withers et al. 2012). To date, all cell wall-precursor modifying ATs belong to a subclade of BAHDs dubbed the “Mitchell clade” (Bartley et al. 2013; Withers et al. 2012). Besides cell wall synthesis proteins, cell-wall-located glycosyl hydrolases (GHs) function in both degradation and growth of cell walls; extracellular

proteins, like arabinogalactan proteins, may function in signaling; and peroxidases catalyze lignin polymerization (Albenne et al. 2013 ; Cosgrove 2016b; Frankova and Fry 2013; Jamet et al. 2006; Passardi et al. 2004). Cell wall enzyme transcript abundance is regulated by transcription factors predominantly in the NAC, MYB, and WRKY families (Gray et al. 2012; Handakumbura and Hazen 2012; Wang and Dixon 2012). In addition to transcriptional regulation, some cell wall synthesis enzymes and transcription factors are regulated by phosphorylation (Chen et al. 2010; Chen et al. 2016; Taylor 2007; Wang et al. 2015b). For example, phosphorylation of Arabidopsis Cellulose Synthase (CES) A1 and CESA3 is important for producing ordered interfaces among cellulose microfibrils or macrofibrils and therefore affects anisotropic elongation of cells (Chen et al. 2010; Chen et al. 2016). In addition, phosphorylation of Arabidopsis CESA7 leads to rapid proteosomal degradation (Taylor 2007). In an example from pine, phosphorylation by PtMAPK6 of a cell wall-related transcription factor, PtMYB4, enhances transcriptional activation (Morse et al. 2009).

Sophisticated manipulation of biosynthetic and regulatory pathways is enabled by an accurate knowledge of the molecules present in a particular biological system. Though more complete transcriptome data for rice, maize, switchgrass, and Setaria stems are available (Bosch et al. 2011; Hirano et al. 2013; Martin et al. 2016; Shen et al. 2013; Zhang et al. 2014), the current grass stem proteome data for rice, Brachypodium distachyon, sugarcanes and bamboo only have a coverage less than 600 proteins (Calderan-Rodrigues et al. 2016; Cui et al. 2012; Douche et al. 2013; Yang et al. 2006). Previous gel-based proteomics studies of rice stems only identified less than 300

proteins, likely missing many proteins involved in important biological processes (Dardick et al. 2007; Nozu et al. 2006). Transcriptome data may be an inaccurate indication of protein abundance and presense due to various post-transcriptional regulatory events (Albenne et al. 2013; Haider and Pal 2013; Vogel and Marcotte 2012; Walley et al. 2016), for example, less than 60% of transcripts are translated to proteins in maize (Walley et al. 2016). There is a similar dearth of metabolite profiling data for grass stems except for sugarcane, which is highly specialized in soluble carbohydrate storage (Glassop et al. 2007). A catalog of rice stem proteins and metabolites will provide a solid foundation for synthetic biology approaches to enhance stem properties and facilitate comparisons with data from other rice organs (Koller et al. 2002) and across species  Kalluri, 2009 #41}.

We report here the abundance of biological components in the second rice internode during elongation, including variation in cell wall components, and an in-depth protein catalog and a low-depth phosphopeptides catalog of the entire internode. In addition, we also report soluble metabolites for entire internodes during elongation and three post elongation stages. We examined expression of putative cell wall metabolism and regulatory proteins in an elongating stem and tentatively identify metabolites enriched in each of the four stem developmental stages. In addition to known cell wall synthesis enzymes that synthesis lignin and polysaccharides, the presence of several putative acyltransferases and glycosyltransferases may indicate their functions in cell wall synthesis. We also detected changes over development in metabolites that have a role in plant-herbivore or plant-pathogen interactions such as tricin 7-glucoside and esculetin.

## Results

During booting when the panicle is just about to emerge from the leaf sheath (BBCH stage 45), internode II of rice stems undergoes rapid elongation and secondary cell wall development. Internode II is the second internode below the panicle, or the internode immediately below the peduncle (Yamaji and Ma 2014). Measurement of rice internode II revealed that it elongates at a relatively constant rate until the panicle fully emerges from the sheath of the flag leaf (Figure 3-1). During elongation of internode II and subsequent panicle maturation, three different datasets were collected on rice internode II during elongation and subsequent panicle maturation. First, we report cell wall composition of eight asymmetrical segments of internode II of elongating stem (ES) during booting. Second, we describe proteomics and phosphoproteomics in the entire internode II at the same stage. Third, we describe metabolites present in entire internode II at ES stage and three later stages, ending at seed maturity.

### *Cell Wall Polysaccharides and Lignin Varies along the Elongating Rice Internode*

Phloroglucinol staining of lignin in internode II of ES was more intense in the mature upper segments than the young lower segments, especially around the vascular bundles (Figure 3-2). We divided the internodes into eight asymmetrical segments (S1 to S8) based on the phloroglucinol staining (Figure 3-2). Segment boundaries were chosen so as to minimize the change in cell wall content within each segment. S1 includes the intercalary meristem but also mature tissue below the meristem (See method). We conducted cell wall analysis on pools of segments from developmentally and physically similar elongating internodes. The analysis revealed lignin, cellulose, and xylose

significant increase along the elongating rice internode, while other components like arabinose and glucose have a decrease or an increase followed by decrease (Figure 3-3).

Acetyl bromide soluble lignin as a fraction of de-starched alcohol insoluble residues (dsAIR) generally increased from the younger to the older internode segments (Figure 3-3A). The change in lignin was not significant from S1 to S4 but dramatically increased from S4 to S8 (Tukey's test, $p<0.05$) (Table 3-S1). The rapid increase in lignin from S4 to S6 was consistent with the phloroglucinol staining from 1.5 to 3.5 cm above the first node.

Cellulose, determined by the anthrone assay after removing hemicellulose, also generally increased from the younger to the older internode segments (Figure 3-3B, ANOVA, $p<0.01$). The mass fraction of cellulose increased continuously from S2 to S7 (Table 3-S1) but did not change significantly between S7 to S8. S1 contained slightly higher cellulose content than S2 but the difference was not statistically significant.

Monosaccharide components of hemicelluloses and other cell wall polysaccharides exhibited different patterns of abundance (Figure 3-3C and D). Except galacturonic acid and glucuronic acid, trifluoroacetic acid (TFA)-released monosaccharides varied significantly across samples (ANOVA, $p<0.01$ for xylose, arabinose, glucose, galactose, and mixed linkage glucan). Xylose and arabinose monosaccharides originate mostly from glucuranoarabinoxylan, the most abundant grass hemicellulose. Xylose increased from S2 to S6 (Tukey's test, $p<0.05$) (Figure 3-3C, Table 3-S1), but did not change

significantly from S1 to S2 and from S6 to S8. Arabinose, did not change significantly from S1 to S3 but decreased continuously from S3 to S8 (Tukey's test, $p<0.05$) (Figure 3-3C). The other major hemicellulose, mixed linkage glucan (MLG), consists entirely of glucose linked via ß-(1-3) and ß-(1-4) bonds. TFA-released glucose mostly comes from MLG, though amorphous cellulose and xyloglucan also contribute. TFA-released glucose peaked at S3 and then gradually decreased. We also measured MLG with a specific lichenase-based assay, which showed a similar, but not identical pattern of abundance compared to the glucose measurement. Lichenase-measured MLG peaked at S4 (Figure 3-3D). TFA-released galactose, mainly from arabinogalactan proteins and pectins, decreased significantly from S2 to S8 (Tukey's test, $p<0.05$) (Figure 3-3C).

Among measured components, cell wall hydroxycinnamoyl esters increased significantly and dramatically along the elongating internode, with a maximum change of 56- and 36-fold for *p*-coumarate (*p*CA) and ferulate (FA), respectively (Figure 3-3E, ANOVA, $p<0.01$). In grass cell walls, these hydroxycinnamic acids are esterified to lignin, arabinoxylan, and possibly other polysaccharides (Buanafina 2009; Lin et al. 2016), with the majority of FA on arabinose residues of glucuranoarabinoxylan, and the majority of *p*CA is esterified to lignin (Molinari et al. 2013). Both hydroxycinnamoyl esters increased from younger to the older internode segments; however, FA and *p*CA showed different patterns of accumulation (Figure 3-3E). FA rapidly increased from S2 to S4, but varied little from S4 to S8. In contrast, *p*CA increased continuously from S2 to S8. The steady increase of the FA:Ara and *p*CA:lignin ratios (Table 3-S1) suggest an increase in both types of cell wall modification across stem development.

As a fraction of alcohol insoluble residue, the total mass of all measured cell wall components except MLG increased from about 500 μg/mg in S1 and S2 segments to about 700 μg/mg in S7 and S8 segments (Figure 3-3F).

***Protein Catalog of Elongating Internode***

To understand the proteins and phosphoproteins available to participate in cell wall changes and other biological processes during stem elongation, we measured trypsin-digested peptides from the whole internode II of ES with three LC-MS/MS experiments (Table 3-1). The same peptide sample was used for experiment 1 and experiment 2. We focus our discussion on experiment 2 conducted by a Waters SYNAPT G2-Si high definition mass spectrometer with ion mobility assisted data-independent analysis ($HDMS^E$) (Waters 2011). Experiment 3 validated the proteins identified in experiment 2 by measuring an independent biological replicate with a Q Exactive Plus mass spectrometer. The phosphoproteomics is conducted on another internode II sample of ES and the identified phosphopeptides were validated by the identifications in the phosphopeptides fraction in experiment 3.

In the whole elongating internode II, experiment 2 identified 2879 proteins and experiment 3 identified 6338 proteins with at least 2 unique peptides. There were 2356 identified proteins shared by experiment 2 and experiment 3 (Table 3-S2). These confidently identified proteins by both experiments correspond to 6% of non-transposon genes in rice. About 53% of the 2356 identified rice proteins were novel identifications,

which have not been detected by at least 2 peptides according to Rice Protoegenomics Database (Figure 3-S1) (Helmy et al. 2012). We therefore have been able to extend the current protein catalog of rice. Subsequent analyses used only the 2356 proteins identified in both experiment 2 and experiment 3.

***Proteome Gene Ontology (GO) and KEGG Ontology (KO) Analysis***

We performed gene ontology (GO) enrichment analysis on cellular component, biological process, molecular function terms for the proteins detected in the elongating rice internode (Table 3-S3).

Cellular Component GO-term enrichment showed some location biases in our dataset (Figure 3-4 and Table 3-S3). Hypergeometric *p*-values indicate that the GO terms of ribosome, cytosol, mitochondrion, cell wall, vacuole, plasma membrane were over-represented. GO terms of endoplasmic reticulum, nucleus, and Golgi apparatus were observed at the expected frequencies. Cell wall synthesis and related proteins localize primarily to the cytosol, Golgi body, plasma membrane, cell wall, and endoplasmic reticulum (Helmy et al. 2012), all of which were reasonably covered here.

Enrichment analysis of biological process and molecular function GO terms revealed some categories overrepresented in the rice elongating stem (Figure 3-5 and Table 3-S3). The most highly overrepresented biological process GO terms were mostly cellular process such as “cellular macromolecule biosynthesis process” and “cellular biosynthetic process”. Metabolic process such as “generation of precursor metabolites

and energy" and "secondary metabolic process" were also significantly over-represented in the elongating internode (Figure 3-5).

Since many metabolic processes were overrepresented in the GO analysis, we conducted KEGG pathway mapping to further examine representation of identified proteins in different metabolic pathways. For all the 2356 proteins identified, we found 945 proteins annotated by 709 KEEG Ontology (KO) terms (Table 3-S4). Among them, 240 of proteins annotated by 156 KO terms were in secondary metabolite biosynthesis pathways such as the phenylpropanoid pathway. The 17 proteins in the phenylpropanoid pathway covered most major enzymes required to synthesize monolignols (Table 3-2, Figure 3-S2), except ferulate 5-hydroxylase and caffeic acid O-methyltransferase.

***Cell Wall Synthesis and Remodeling Proteins, and Extracellular Proteins***

From the list of identified proteins, we highlight proteins from our dataset that might be involved in the changes in cell wall composition observed along the elongating internode. We examined the GT2, GT43, GT47, GT61 and AT clades that contain known grass cell wall synthesis proteins. We identified two GT2, four GT61, and six Mitchell clade-AT proteins (Table 3-3). OsIRX10, a GT47 involved in xylan synthesis (Chen et al. 2013), though not identified in experiment 3, is identified by 2 unique peptides in experiment 2. Though no GT43s were identified by two unique peptides, OsGT43A (LOC_Os05 g03174), involved in xylan synthesis (Lee et al. 2014a), was identified by one unique peptide. The ATs identified consist of four hydroxycinnamoyl CoA transferases implicated by biochemical and genetic studies in cell wall

modification and two putative feruloyl CoA transferases that have not been described (Table 3-3).

To identify grass cell wall remodeling genes, we examined 18 GH families that either modify or degrade specific plant cell wall components (Minic 2008), have been found in rice cell walls (San Clemente and Jamet 2015), or belong to a monocot-expanded family (Sharma et al. 2013). We identified 39 proteins in these families (Figure 3-6 and Table 3-S5). In the GH3, GH16, GH17, GH28 and GH31 families, more than 3 proteins were identified (Figure 3-6).

Many extracellular proteins identified by cell wall proteomics had been collected in WallProtDB (Chen et al. 2009; Cho et al. 2009; Jung et al. 2008). We found 43 proteins in WallProtDB present in the elongating internode. The two most abundant categories were GHs and cell wall-localized type III peroxidases that may catalyze polymerization of monolignonols (Passardi et al. 2004) (Figure 3-S3 and Table 3-S6). Besides these two major groups, we also identified eight proteases, three leucine-rich repeat proteins that may participate in plant defense, two expansins that may function in cell wall-loosening, and two fasciclin-like arabinogalactan proteins implicated in plant development and stress-response (Cosgrove 2000; Jamet et al. 2006; Johnson et al. 2003).

***Kinases and Transcription Factors***

Cell wall biosynthesis and other stem developmental processes are regulated by numerous transcription factors and kinases (Handakumbura and Hazen 2012; Wang et al. 2015b). From the rice kinase database (Dardick et al. 2007) and rice TF database, we identified 29 protein kinases including both receptor and cytoplasmic kinases (Table 3-S7) and 44 proteins as transcription factors (Table 3-S8).

We categorized protein kinases based on phylogeny, presence of a conserved kinase motif, and predicted subcellular localization (Figure 3-7). The literature reports protein kinases with cell wall phenotypes or that regulate cell wall genes belonging to three groups: TKL (Tyr-kinase-like) kinases, CMGC (CDK, MAPK, GSK3, and CLK) kinases, and CAMK (calcium/calmodulin-dependent) kinases (Matschi et al. 2013; Morse et al. 2009; Oh et al. 2011). We found evidence of the presence of 14 TKL kinases, including four from the LRR receptor-like kinase family. The six CAM kinases identified were all calcium-dependent kinases. Two CMGC kinases identified were both from the MAPK family. We also checked if the identified kinases contain a conserved arginine in the RD-motif in kinase subdomain VI. The absence of this motif (i.e., non-RD) correlates with functioning in pathogen response (Dardick and Ronald 2006). We found 19 RD kinases and only 4 non-RD kinases (Figure 3-7B). Kinases that regulate plant development and cell wall biosynthesis are often localized to the plasma membrane (Hematy et al. 2007; Oh et al. 2011; Park et al. 2001). Five of the 29 identified proteins were predicted to be in the secretory pathway and could be plasma membrane- localized (Figure 3-8C).

As phosphorylation mediated by protein kinases is an important regulatory mechanism for plant development and cell wall synthesis (Hematy et al. 2007; Oh et al. 2011; Park et al. 2001), we conducted a small phosphoproteomics experiment to explore phosphoproteins in the elongating internode. The identified phosphopeptides were further confirmed by their presence in the phosphoprotein enriched fraction in experiment 3. We consistently identified 21 unique phosphopeptides from 20 phosphoproteins (Table 3-S9). Among them, 9 were phosphoproteins not reported previously in P3DB, including an LRR-III family receptor like kinase (LOC_Os03g12250).

We also identified 41 protein transcription factors. The C3H family was most common (Figure 3-S4). We also identified a few proteins in cell wall-related or cell division-related transcription factor families: CDC5 (LOC_Os04g28090) from the MYB related family and NAC2 (LOC_Os08g06140) from the NAC family (Feller et al. 2011; Wang et al. 2015a).

***Metabolite Profiles of the Elongating Internode, Mature Internodes, Leaf and Root***

The proteomics data confirmed the importance of secondary metabolic processes in the internode II of ES, thus, we also analyzed methanol-soluble secondary metabolites, including phenylpropanoids, isolated from different developmental stages of rice internode II. We collected internode II at the following four stages: the same elongating stem (ES) internode as the proteomics experiment, the early mature stem (EMS)

internode at flowering, the mature stem (MS) internode at grain filling, and the post mature stem (PMS) internode after seed maturity (Figure 3-6). For comparison, we also extracted metabolites from rice leaves and roots. We analyzed three biological replicates by LC-MS/MS with technical triplication in both negative and positive ionization modes. We identified 11,929 and 8,521 metabolite ions, respectively, that were reproducibly detected among replicates of any sample (See Methods; Table 3-S10 and Table 3-S11). Of these, we tentatively annotated 22 negative ions and 53 positive ions based on mass similarity, isotope similarity, and theoretical fragmentation. We used commercial standards to confirm the MS/MS pattern and retention time of a phenylpropanoid, *p*CA, and a flavonoid, apigenin.

Most metabolite ions varied significantly among organs and across stem development. As expected, principal component analysis (PCA) showed clear metabolic differences between stem metabolite profiles and those of roots and leaves (Figure 3-8A & B). PCA of just the ions from stem samples mostly separated the different stages from each other, consistent with differences in metabolite profiles during stem development (Figure 3-8C & D). Indeed, 6338 negative metabolite ions and 3110 positive metabolite ions varied significantly across stem development (ANOVA, $q$-value<0.01; Table 3-S10 and Table 3-S11). The ions verified with authentic standards, *p*CA and apigen, varied significantly among organs though not across internode stages (Figure 3-S5). K-means clustering of the metabolite dataset z-scores arranged in developmental order show that the metabolites are well-described by five clusters, with similar cluster patterns in the positive and negative ionization modes (Figure 3-9). The five clusters

correspond to metabolites that predominate in each of the four developmental stages and another cluster that is represented in both the ES and EMS samples. Table 3-4 lists the number of ions identified in each ionization mode for each cluster and some tentatively identified metabolites, with additional identifications listed in Table 3-S10 and Table 3-S11. For example, cluster 1 metabolites were most abundant in ES. This cluster in negative ion mode contained 1021 metabolites including coniferyl aldehyde. Cluster 1 in positive ion mode contained 667 metabolites including methyl cinnamate.

By examining the Plant Metabolic Network, we also found some enzymes related to the identified metabolites present in internode II of ES according to our proteome data. For example, C4H (LOC_Os05g25640), which synthesizes *p*CA from cinnamic acid and CCR (LOC_Os08g34280), which synthesizes coniferyl aldehyde from Feruloyl-CoA both present in ES and their direct products were detected. Two putative tricin synthases (LOC_Os08g38900, LOC_Os08g38910) that synthesize tricin from tricetin, which is synthesized from apigenin, were also present. Their presence together with the presence of apigenin and tricin 7-glucoside, a tricin derivative, suggest the tricin synthesis pathway is active.

## Discussion

The basipetal development of the grass internode renders it an excellent system for gaining molecular understanding of secondary development in monocots. In the internode, different developmental stages are available simultaneously under highly similar environmental conditions. Complementing recent transcriptome studies (Bosch et al. 2011; Hirano et al. 2013; Martin et al. 2016; Shen et al. 2013; Zhang et al. 2014), this study contributes novel grass stem proteomics, phosphoproteomics, and metabolite datasets toward understanding grass stem development and functions, specifically for the elongating rice internode.

### *Cell Wall Changes Associated with Stem Development*

The cell wall measurements of rice elongating internode segments are mostly consistent with similar measurements from maize (Zhang et al. 2014). The patterns of lignin, cellulose and xylose were similar to those measured by methylation-based linkage analysis in maize elongating internodes. However, MLG and TFA-released glucose showed an initial increase and then decrease in rice elongating internode in contrast to the monotonic decrease of MLG in the maize elongating internode (Figure 3-3). MLG accumulation in maize and rice may be different, or alternatively, the short length of our segments S1 to S4 may have permitted us to detect this variation in young cell walls, missed in the uniform 1 cm segments of Zhang et al. The general consistency of rice and maize internode cell wall profiles during development suggests that the rice proteomics data can be used to gain insight into cell wall development of other grasses.

We extended the previous internode cell wall profiling with measurements of hydroxycinnamoyl esters, which play an important role in cross-linking cell wall polymers in grasses (Buanafina 2009). Both *p*CA and FA increased from young segments to old segments but showed different patterns of change, with *p*CA amounts continuing to increase through development but FA remaining constant in old segments. When integrated with gene expression or proteomics data of these segments, the difference in *p*CA and FA accumulation may facilitate identification of the transferases and other proteins that function in their incorporation, a topic of considerable recent interest (Buanafina et al. 2016; Chateigner-Boutin et al. 2016; Lin et al. 2016; Molinari et al. 2013).

***Cell Wall Synthesis, Remodeling, and Protein Phosphorylation are Important During Stem Elongation***

This study applied different proteomics methods to unveil the protein catalog in an elongating rice internode and provided technical guidance for future proteomics of internodes. Though many known or hypothesized cell wall-related proteins were identified (Table 3-2, Table 3-3, and Figure 3-6), others were absent from the dataset, especially some transcription factors from MYB and NAC family that controls cell wall synthesis. Coverage in future experiments can be improved by conducting further fractionation to reduce the complexity of protein samples, such as subdividing the sample, subcellular fractionation, or additional LC separation or using the combinatorial peptide ligand library (Li et al. 2013; Righetti and Boschetti 2016; Wang et al. 2010a).

Within the whole elongating rice internodes actively undergoing cell wall alterations, we detected 2356 proteins, including several known or implicated in cell wall synthesis. These observations reinforce the importance of the identified proteins. Since we detected them by LC-MS/MS with only minimal fractionation, these proteins are likely to be the most abundant representatives of their corresponding families, when family members can be distinguished. We detected proteins corresponding to almost the entire monolignol biosynthesis pathway (Table 3-2) (Humphreys and Chapple 2002). We detected one member in PAL, C4H, C3H, CCoAOMT, and CCR families, and two members in CAD, HCT and 4CL families. The one or two enzymes detected at each step of phenylpropanoid pathway could be the major enzymes that catalyze monolignol biosynthesis in rice stem. We also detected 5 peroxidases that are good candidates for acting in monolignol polymerization. One of the identified GT2s is CESA1, a likely primary cell wall cellulose synthase in rice (Table 3-3) (Wang et al. 2010b). Though only from experiment 2, we also detected OsIRX10, a GT47 involved in stem xylan synthesis. Mutants of this gene exhibit shorter stems, thinner secondary cell walls, and decreased cell wall xylose content (Chen et al. 2013). A GT61 we identified, OsXAX1, participates in synthesis of a xylan side chain (Chiniquy et al. 2012). Rice *xax1* mutants have decreased FA and xylose. Among the six identified ATs, AT10 and AT4 have been previously reported to play a role in rice cell wall synthesis. AT10 incorporates *p*CA into arabinoxylan (Bartley et al. 2013), and OsAT4 functions as a *p*-Coumaroyl-CoA:monolignol transferase that involve in lignin synthesis (Petrik et al. 2014; Withers et al. 2012). Functional data are also available for *Brachypodium* orthologs of rice OsAT3 and OsAT1 (Buanafina et al. 2016; Petrik et al. 2014). OsAT2 and OsAT9 have

not been characterized but are candidate feruloyl transferases based on expression patterns and a phylogenetic study in *Brachypodium* (Molinari et al. 2013). The correlation between AT9 and FA abundance in rice above-ground tissues also support its role as a feruloyl transferase (Lin et al. 2016)

Many GHs we identified in the internode proteomics dataset are from families that may function in cell wall remodeling and defense responses (Figure 3-6) (Minic 2008). GH3 and GH51 enzymes may remodel arabinoxylans. GH3s from barley seedlings and Arabidopsis stems have arabinofuranosidase or xylosidase activity (Lee et al. 2003; Minic et al. 2004). Two GH51s from barley have arabinofuranosidase activity on arabinoxylan (Ferre et al. 2000; Lee et al. 2001). These enzymes might be partially responsible for the decrease in arabinose:xylose ratio across internode development (Table 3-S1). Other families may function in minor polysaccharides in grass cell walls or pathogen cell walls. For example, the GH27 protein we detect may has galactosidase activity (Kim et al. 2002) and responsible for the decrease of galactose from young to old stem segments (Figure 3-3), though we cannot rule out that the decreases are caused simply by dilution due to addition of other components to the wall. Some rice GH16 enzymes exhibit xyloglucan endotransglucosylase and xyloglucan endohydrolase activities (Hara et al. 2014). GH17 enyzmes have been found to be endo-1,3-β-glucosidases that degrade pathogen cell walls, functioning in defense (Minic 2008); while, GH38s have mannosidase activity in bacteria, but unknown functions in plants (Suits et al. 2010).

We identified many protein kinases and novel phosphopeptides in the elongating internode, consistent with a role for protein phosphorylation in controlling stem development (Table 8). Currently, protein kinases with cell wall phenotypes or that regulate cell wall genes are mainly in three kinase groups. These include the LRR (Leucine-rich repeat) receptor like kinases, such as THE1 and FEI1 from Tyr-kinase-like (TKL) group; MAPK6, which is in the CMGC (CDK, MAPK, GSK3, and CLK) group; and CDK28, from the CAMK (calcium/calmodulindependent protein kinase) group (Matschi et al. 2013; Morse et al. 2009; Oh et al. 2011). More than half of kinases identified in this study are from these three groups. One of the kinase from TKL group that we identified is a putative ortholog of the BRASSINOSTEROID INSENSITIVE 1-ASSOCIATED RECEPTOR KINASE 1 (BAK1, LOC_Os02g09359). In Arabidopsis, BAK1 can form a heterodimer with BRASSINOSTEROID INSENSITIVE 1 (BRI1) in the presence of brassinosteriod to regulate plant development including inflorescence stem growth and secondary cell wall formation (Nam and Li 2002).

Non-RD kinases, i.e., those that lack a conserved RD in the kinase subdomain IV, are predicted to function in pathogen recognition, while RD kinases are likely to be involved in other biological processes (Dardick et al. 2007; Dardick and Ronald 2006). The under representation of non-RD kinases in the elongating stem is consistent with pathogen detection not being a major function in the stem under our growth conditions, though could also be due to low expression of non-RD kinases relative to RD-kinases.

In our data, there are 5 identified kinases are likely to be processed through the secretory system. Many membrane-associated kinases, like WAK1, BRI1, and THE1, impact cell wall formation (Hematy et al. 2007; Oh et al. 2011; Park et al. 2001). We also identified a novel phosphopeptide from a membrane-associated LRR-III family receptor-like kinase. Autophosphorylation of LRR receptor like kinases is important for activation of plant growth regulation, disease resistance, and stress response signaling pathways (Mitra et al. 2015; Tor et al. 2009).

***Secondary Metabolites Vary During Stem Maturation***

During stem development, we observe that many metabolites change significantly. The vast diversity of secondary metabolites in plants evolved as protection against pathogens, insects, and animals (Dixon 2001; Gershenzon and Dudareva 2007; War et al. 2012), including functioning indirectly by alerting predatory insects to herbivory and other chemical ecological effects (Schuman and Baldwin 2016). Abundance of some rice pathogen-repelling metabolites varies across development. For example, diterpenoid phytoalexins, which repel rice blast fungus, are less abundant in young rice leaves relative to old leaves (Kodama et al. 1988). Susceptibility to insects like stemborers also differs during rice development (Bandong and Litsinger 2005; Viajante and Heinrichs 1987 ). For example, rice is susceptible to yellow stemborer at flowering but not at pre-booting or after panicle emergence (Viajante and Heinrichs 1987).

Correlation between rice metabolite abundance and pathogen and pest resistance across development might indicate a role for the metabolites in resistance (Gardner 1977; War et al. 2012). For example, 6,7-dihydroxycoumarin (esculetin) and phenylacetaldehyde were more abundant in the elongation and early mature stems (ES and EMS) but less abundant in the mature and post mature stems (MS and PMS). Esculetin has fungitoxicity and may be involved in defense (Gomez-Vasquez et al. 2004). On the other hand, tricin 7-glucoside, a phagostimulant that triggers feeding in insects like the small brown planthopper (Adjei-Afriyie et al. 2000; Bouaziz et al. 2001), remained relatively low until the post-mature stage, when grains are fully mature.

This study catalogs cell wall changes, protein abundance, and secondary metabolite profiles of rice elongating internodes. The novel and relatively deep proteomics data reveals the presence of metabolic enzymes, including those for cell wall synthesis and remodeling, in agreement with the dramatic changes we observe in cell wall components. In addition, the changes of metabolite profiles associated with stem maturation could indicate potential defensive roles of secondary metabolites during plant development. The availability of the protein catalog and metabolite profiles provides a crucial tool for understanding fundamental molecular processes of grass stem development, over inferences from transcriptomics alone. These data will facilitate identification of protein targets for optimizing stem development and cell wall composition of grasses to improve their agronomic properties and downstream uses.

## Materials and Methods

### Rice Growth Conditions, and Material Harvest

Following an adaptation of a published protocol (Eddy et al. 2016), *Oryza sativa* ssp. *japonica* cv. *Kitaake* seedlings were grown in a mixture of Turface Athletics medium:vermiculite (1:1) in a greenhouse at 29-32°C during the day and 24-25°C during the night. Natural day lengths less than 13 h were supplemented with artificial lighting. Two weeks after germination, plants were fertilized three times per week with JACKS PROFESSIONAL LX 15-5-15 4CA2MG fertilizer.

Plant samples were collected at development stages defined by "Biologische Bundesanstalt, Bundessortenamt und CHemische Industrie" (BBCH) identification keys of rice (Lancashire et al. 1991). Stem internodes II (the second internode from the top) at booting stage (BBCH stage 45) were harvested as the elongating stem internode (ES) for cell wall, proteomics, and soluble metabolite profiling. At this stage, the top of panicle is 0-1 cm beneath the top of flag leaf sheath. Internode II were cut right above node II where the leaf sheath and axillary bud attached to stem and at the dark green ring below node I. Axillary buds are carefully removed. For metabolite profiling, additional internode II at flowering stage (BBCH stage 65), grain filling stage (BBCH stage 75) and seed mature stage (BBCH stage 92) were harvested as early mature stem internode (EMS), mature stem internode (MS), post mature stem internode (PMS). The flag leaf blade and all roots of a plant in grain filling stage were harvested as mature leaf and root samples for metabolite profiling. Root samples were washed with de-ionized water for 5 min. All samples for cell wall assays, proteomics and metabolite

profiling were immediately frozen in liquid nitrogen and stored at -80ºC.

For cell wall assays, twenty internode II at ES with an average length of 10 cm were unevenly divided to segments and pooled per biological replicate, with 3 to 4 biological replicates collected, depending on the assay. The eight uneven segments from the base of internode II were as follows: S1-S4, the first four successive 5 mm segments; S5-S7, the next three 10 mm segments; and S8, the remainder of the internode. The uneven sampling strategy was designed based on the phloroglucinol staining data, which revealed rapid changes in lignin near the base of the internode, consistent with previously observed rapidly changing cell wall content in the basal part of the maize internode (Zhang et al. 2014). S1 segment locates right above the node II. Previous study haven shown that the intercalary meristem is about 2 mm above the node (Kende et al. 1998) The tissue below meristem exist before internode elongation and therefore is mature tissue.

For proteomics, six developmentally matched internode II of ES were ground with a mortar and pestle with liquid nitrogen. This fine-ground stem material was divided into six equal parts with 3 technical replicates for Gel-LC-MS/MS (in-gel tryptic digestion followed by liquid chromatography-tandem mass spectrometry) and 3 technical replicates for in-solution digestion followed by liquid chromatography-tandem mass spectrometry.

For metabolite profiling, all samples were pooled from two plants per biological replicate and three biological replicates were collected.

**Microscopy**

We performed phloroglucinol staining on cross-sections of the internode II of ES as previously described (Liljegren 2010). Internode sections were stained with 50 µL phloroglucinol solution for 2 min on slides and examined immediately after adding 50 µL 50% (v/v) HCl. The sections were observed with a Leitz Dialux 20 microscope with a Leitz Wezlar 10x lens and imaged with an Olympus DP71 camera on automatic exposure mode controlled by DP controller software (ver 3.1.1.267).

**Cell Wall Measurements and Data Analysis**

For all cell wall assays, fresh-frozen internode II segments (S1 to S8) were ground, made into alcohol insoluble residue (AIR), destarched, and assayed for cell wall composition as described previously (Bartley et al. 2013). Briefly, hemicellulosic and pectic monosaccharides were released with 2 M TFA at 120°C for 2 hrs and measured by high performance ion exchange chromatography. The TFA-insoluble pellet was used for cellulose measurements via an anthrone assay. Lignin was measured by solubilization with 25% acetyl bromide in glacial acetic acid followed by absorbance measurements at 280 nm with a micro-plate reader. HCAs were released by incubation with 2 M NaOH at 25 °C for 24 h and measured by a high performance liquid chromatography with a UV detector (Bartley et al. 2013). MLG was measured with a lichenase-based kit (Megazyme, K-BGLU) as described previously (Vega-Sanchez et al. 2012). Three biological replicates were used for all experiments expect the lignin assay, which used four biological replicates. The R general package was used for data analysis.

**Protein Extraction and Sample Preparation**

For experiment 1 and experiment 2, total proteins were extracted from each technical replicate with a phenol-based method previously described (Lee et al. 2010; Saravanan and Rose 2004). Briefly, plant material was re-suspended with extraction buffer consisting of 0.1 M Tris-HCl (pH 8.0), 10 mM EDTA, 0.9 M sucrose, 0.1% DTT, 1% protease inhibitor cocktail (Sigma P9599). Proteins were extracted with an equal volume of Tris buffered phenol (pH 8.8) for 30 min at 4ºC, and then precipitated and washed three times with 0.1 M ammonium acetate. The protein pellets were dried at room temperature and solubilized with 7 M urea, 50 mM Tris-HCl, pH 8.0. Protein samples were quantified with a BioRad Protein Assay Kit using BSA as a standard. For peptide preparation, 150 μg proteins of each technical replicate were loaded on three lanes on a 12% (w/v) polyacrylamide gels with 5% (w/v) stacking gels. Sodium dodecyl sulfate polyacrylamide gel electrophoresis was conducted on a BioRAD gel Cryterion$^{TM}$ apparatus at an initial voltage of 120 V for 10 min followed by 100 V constant voltage in 1X Tris-Glycine running buffer. The gel was stained by Coomassie brilliant blue and divided into 5 slices corresponding to apparent molecular mass. Each gel slice was handled separately and diced into 1 mm cubes and incubated in detaining solution (200 mM ammonium bicarbonate and 40% acetonitrile) for 30 min twice at 37ºC. The distained gel slices were dried and digested with proteomics-grade trypsin (Sigma, T6567). All peptide samples were equally split for LC-MS/MS in experiment 1 and experiment 2.

For experiment 3, ground internode tissue was divided to four 15 mL centrifuge tubes

(1.5 g/tube) containing 15% (w/w) Polyvinylpolypyrrolidone (PVPP) as four technical replicates. To each tube, 10 mL of 0.1M ammonium acetate in MeOH with 2% β-mercaptoethanol was added and incubated at -20 °C for 2 h. Samples were centrifuged at 4500 x g at 4°C for 10 min and the supernatant was discarded. The wash step was repeated 3 times, with the incubation step included between washes, as described above. After removing the supernatant from the final wash, pellets were dried using a Turbo Vap (Biotage, Charlotte, NC) under a stream of nitrogen for 10 min. 4 mL of a protein solubilization solution containing 7M urea, 2M thiourea, 4% 3-[(3-cholamidopropyl) dimethylammonio]-1-propanesulfonic, 10 mM tris(2-carboxyethyl)phosphine in 100 mM $NH_4HCO_3$ was added and the samples were stored overnight at 4°C, then sonicated for 1 min in a sonication bath (Branson, Danbury, CT). Following sonication, samples were incubated for 30 min at 60°C with shaking at 1400 rpm, then briefly sonicated, vortexed, and centrifuged at 4500 x g, 4°C for 10 min. Prior to digestion, a Coomassie assay was performed on the supernatants containing the proteins. For digestion (with alkylation using chloroacetemide, 5 mM final concentration), the supernatants from each tube were transferred to 50 mL centrifuge tubes and diluted 10-fold with 50 mM $NH_4HCO_3$ and 100 μL of 1M $CaCl_2$ was added. 10 mg/mL sequencing grade, modified trypsin was added in a 1:500 (w/w) trypsin:protein ratio. The samples were incubated at 37°C for 3 h with gentle shaking. Digested proteins were desalted and washed using 100 mg solid-phase extraction strong cation exchange (SPE-SCX) columns (Discovery DSC-SCX, Sigma-Aldrich, St. Louis, MO) on a GX-274 Liquid Handler (Gilson, Middleton, WI). Samples were concentrated using a SpeedVac SC 250 Express (Thermo Scientific, Waltham, MA). A second SPE-SCX wash step was performed

using a vacuum manifold (Visiprep, Sigma-Aldrich, St. Louis, MO). Samples were concentrated in a SpeedVac and a Bicinchoninic Acid Protein Assay (BCA) (Smith et al. 1985) assay was performed. Peptides were stored at -70°C.

Peptides (200 µg) from each of the 4 technical replicates were dried in a SpeedVac. 30 µL of 1 M triethylammonium bicarbonate (TEAB) was added to the dried peptides and 4-plex iTRAQ™ labeling (Abdi et al. 2006) was performed following the manufacturer's instructions (SCIEX, Framingham MA). 300 µL of water was added to hydrolyze the samples prior to combining the labeled samples together. The combined samples were concentrated and washed using SPE C18. The eluted sample was concentrated and a BCA assay was performed.

A high-pH reversed phase liquid chromatography separation (HPRPLC) was performed using an Agilent 1100 HPLC System (Agilent, Palo Alto, CA) equipped with a quaternary pump, degasser, diode array detector, peltier-cooled autosampler, and fraction collector (set to 4°C for all samples). The separation columns consisted of an XBridge C18 250 x 4.6 mm analytical column containing 5 µM particles and equipped with a 20 x 4.6 mm guard column (Waters, Milford, MA) (Wang et al. 2011). Ninety-six fractions were collected into microwells, with the first fraction collected after 15 minutes into the gradient. Fractions were combined (rows) into 12 fractions, then concentrated using a SpeedVac.

**Mass Spectrometry and Peptide Identification for Experiment 1**

The LC-LTQ Orbitrap XL was run at the University of Missouri, Proteomics Center. Samples were resuspended in 21 µL aqueous solution with 5% acetonitrile and 1% formic acid. Samples were vortexed, centrifuged, and transferred to autosampler vials. The vials were placed in a thermostatted (7°C) autosampler in the Proxeon Easy nLC system. A full loop injection (18µL) of each sample was loaded onto a C8 trap column (Pepmap100 C8 trap column Dionex/Thermo Scientific). Bound peptides were eluted from this trap column onto a 25 cm x 150 µm inner diameter pulled-needle analytical column packed with HxSIL C18 reversed phase resin (5 µm particle size, 100 Å pore size, Hamilton Co). Peptides were separated and eluted from the analytical column with a step gradient of solvent A (0.1% formic acid in water) and solvent B (99.9% acetonitrile, 0.1% formic acid) at 400nL/min at room temperature. Initial conditions were 5% B, and ramping to 20% B from 2 min to 20min, then ramping to 30% B at 57 min, then ramping to 90% B at 62 min, hold at 90% B from 62 min to 84 min, back to 5% at 85 min and hold at initial conditions until 90 min.

For the LC-LTQ Obitrap XL spectrometer (Thermo Fisher Scientific Inc., Waltham, MA, USA), nanoElectrospray positive-ion was used. FTMS data were collected at a resolution of 30000, 1 microscan, 300-1800 m/z, profile and a cycle of approximately 3 seconds. The 9-most-abundant peptides were selected for ion-trap collision-induced dissociation MS/MS (2 m/z mass window, 35% normalized collision energy, centroid). Dynamic exclusion was enabled with the following parameters: repeat count 1, repeat duration 30 seconds, exclusion list 500, and exclusion duration 180 seconds.

The results of LC-LTQ Obitrap XL were analyzed by Proteome Discoverer (version 1.4, Thermo Scientific) and Scaffold. The rice protein sequence FASTA file was imported and indexed for searching in Proteome Discoverer. A SEQUEST HT search and filtering was performed with the following parameters: tryptic peptides with a maximum mis-cleavage of 2, mass range 350 to 5000 Da, peptide mass tolerance 25 ppm, fragment mass tolerance 0.8 Da, carbamidomethyl Cys and Met oxidation as variable mods. Data were then exported to Scaffold and filtered with peptide mass tolerance 10 ppm, peptide confidence > 95%, peptide per protein ≥ 2, and protein confidence > 99%, peptide FDR < 5.1%. Search results were exported to Scaffold and the results from different gel-slices were merged.

**Mass Spectrometry and Peptide Identification for Experiment 2**

For the proteomics with the LC-HDMS$^E$ method, peptide samples were resuspended in 50 µL of 3% acetonitrile with 0.1% TFA and 5 fmol/µL yeast alcohol dehydrogenase (ADH1_YEAST, SwissProt P00330) as an exogenous standard for relative protein quantitation. Variable injection volume was used for each sample to maximize protein loads to 100 ng by adjusting the concentration of ADH1 to a final concentration of 1 fmol/µL based on the concentration determined by scouting runs. A Symmetry C18 column (180 µm x 20 mm) was used as trap column and a C18 HSS T3 column (75 µm x 150 mm) was used as the analytical column (Waters Corporation, Milford, MA). Nanoflow UPLC was performed at a flow rate of 500 nL/min. The LC gradient consisted of solvent A (0.1% formic acid in water) and solvent B (99.9% acetonitrile,

0.1% formic acid) starting at 5% B and ramping to 35% B over 90 min, then ramping to 50% B to 95 min, then ramping to 85% B at 97 min, back to 5% at 107 min and re-equilibrated until 120 min. Each sample was run in triplicate for statistical evaluation of technical reproducibility.

A hybrid ion mobility quadrupole time-of-flight mass spectrometer, SYNAPT G2-S*i* HDMS (Waters, Wilmslow, UK) was used to identify and quantify the relative abundances of the tryptic peptides. LC-MS/MS experiments were acquired using ion mobility assisted data-independent analysis acquisition mode similar to that previously described (Distler et al. 2014). Data were collected at 25000 resolution, 50-2000 m/z, 0.8 second/scan. The Transfer T-wave fragmentation method that we used applies a linear collision energy ramp from 17-60 V along the 20-200 drift bin scale (total of 200 drift bins) for the elution of precursor peptide ions.

The results of the LC-HDMS$^E$ experiment were analyzed with Progenesis™ QIP v2.0 for proteomics (Waters, Wilmslow, UK). Each experiment was imported and chromatographically aligned. The peptide was identified by a previously described ion accounting algorithm (Li et al. 2009) with rice protein sequences download from Rice Genome Annotation Project (Release 7, http://rice.plantbiology.msu.edu/index.shtml) (Ouyang et al. 2007). The search parameters were: tryptic peptides with a maximum mis-cleavage of 1, peptide mass tolerance 20 ppm, peptide accounting score >5, max protein mass 250 kDa, default protein modification (carbamidomethyl Cys and Met oxidation as variable mods), fragments per peptide ≥ 3, fragments per protein ≥ 7,

peptides per protein ≥ 1, FDR < 4%. Proteins identified by at least 2 unique peptides were analyzed. Proteins were quantified by the average of the 5 most abundant peptides or of all peptides if the number of peptides was less than 5 (Silva et al. 2006). The absolute amounts of proteins with no less than two unique peptides were calculated based on the response factors (counts per fmol) of an alcohol dehydrogenase standard of each run.

**Mass Spectrometry and Peptide Identification for Experiment 3**

Peptides were separated prior to MS analysis using a Waters nano-Acquity UPLC system system (Waters, Milford MA) configured for on-line trapping of a 5 µl injection (0.1 µg/µl peptide concentration) at 3 µl/min with reverse direction elution onto the analytical column at 300 nL/min. Separation occurred by way of in-house packed columns using Jupiter C18 media (Phenomenex, Torrence, CA), 5 µm particle size for trapping column (100 µm x 4 cm length) and 3 µm particle size for the analytical column (75 µm i.d. x 70 cm length) coupled to HF etched fused silica tips (Kelly et al. 2006). Mobile phases consisted of (A) 0.1% formic acid in water and (B) 0.1% formic acid in acetonitrile with the following gradient profile (min, %B): 0, 1; 2, 8; 20, 12; 75, 30; 97, 45; 100, 95; 110, 95; 115, 1; 150, 1.

Tandem mass spectra (MS/MS) were generated for each fraction using high-energy collision dissociation (HCD) coupled to Q Exactive Plus mass spectrometer (Thermo Scientific, San Jose, CA) outfitted with a home made nano-electrospray ionization interface. The ion transfer tube temperature and spray voltage were 325ºC and 2.2 kV, respectively. Data were collected for 100 min following a 15 min delay from sample injection. FT-MS spectra were acquired from 400-2000 m/z at a resolution of 35k

(Automatic gain control target 3 x $10^6$) and while the top 10 FT-HCD-MS/MS spectra were acquired in data dependent mode with an isolation window of 2.0 m/z and at a resolution of 17.5k (Automatic gain control target 1 x $10^5$) using a normalized collision energy of 30 and a 30 s exclusion time.

Spectra were converted to .dta files using Bioworks Cluster 3.2 (Thermo Fisher Scientific, Cambridge, MA) and amino acid sequences assigned to tandem mass spectra using the MS-GF+ search algorithm (Kim and Pevzner 2014) against the IRGSP-1.0_protein.fasta (2016) database for *O. sativa*. Search parameters consisted of a 10 ppm tolerance for precursor ion masses, and a ± 0.5 Da window on fragmentation masses, allowed dynamic modifications included oxidation of methionine (15.9949 Da), and static modifications included IAA alkylation of cysteine (57.0215 Da), and iTRAQ modification of lysine and N termini (+144.1021). Missed cleavages were allowed for full and partial tryptic peptides. Peptides were filtered to achieve a false discovery rate ≤ 1% using spectral probability values generated by the search algorithm. A protein was considered positively observed if 2 or more unique peptides were measured for it across all iTRAQ labeled replicates. The ID of identified proteins are converted to Rice Genome Annotation Project ID by the online converter at OryzaExpress: http://bioinf.mind.meiji.ac.jp/OryzaExpress/ID_converter.php

**Sample preparation and Mass Spectrometry for Phosphoproteomics**

Similarly for phosphoproteomic analysis, six developmentally matching internode II of ES were ground with liquid nitrogen, divided into 3 replicates each for in-gel digestions. Proteins were extracted and digested with the same methods as experiment 1 and experiment 2 but with the addition of 1% Phosphatase Inhibitors cocktails 1 & 2

(Sigma P2750 and Sigma P5726). For in gel digestion, 500 μg of total proteins were loaded to three lanes for sodium dodecyl sulfate polyacrylamide gel electrophoresis. Each lane on the gel was excised, subjected to in-gel trypsin digestion and then pooled. Tryptic digested peptides were enriched for phosphopeptides using Pierce™ TiO2 Phosphopeptide Enrichment and Clean-up Kit (Thermo Fisher Scientific, 88301 and 88303) (Larsen et al. 2005). The eluted phosphopeptides were acidified with 2.5% TFA to pH 2.0 to 2.5, and cleaned with graphite column, lyophilized, and stored at −80°C.

Phosphoprotein data were generated under contract by Pepproanalytics, LLC. Lyophilized tryptic phosphopeptides were resuspended in 25 μL of 0.1% formic acid and 10 μL was loaded. Peptides were desalted using C8 Captraps (Bruker-Michrom Bioresources, Auburn, CA). Chromatographic separations were achieved using a 55 min linear acetonitrile gradient (5%-35% acetonitrile, 0.1% formic acid) with a column packed with "Magic C18" (200 Å, 5 μm bead, Bruker- Michrom Bioresources) stationary phase. The phosphoproteomics was conducted by LTQ Orbitrap XL with electron-transfer dissociation (Thermo Fisher Scientific, San Jose, USA). Survey scans (MS1) used the following settings: analyzer: FTMS; mass range: normal; resolution: 60,000; scan type: positive; data type: centroid; scan ranges: 200-2000 m/z. The top 7 masses from the survey scans were selected for data dependent acquisitions. Data dependent scan settings were the following: analyzer: ion trap; mass range: normal; scan rate: normal; data type: centroid. Dynamic exclusion, charge state screening and monoisotopic precursor selection were enabled. Unassigned charge states and masses with a charge state of one were not analyzed. The "data dependent decision tree" option

was enabled as previously described (Swaney et al. 2008), allowing for fragmentation of peptides using collision induced dissociation or electron transfer dissociation.

The phosphoprotoemics results were analyzed with Proteome Discoverer (version 1.3, Thermo Scientific). Acquired spectra were searched against the rice protein sequence FASTA file concatenated with a randomized decoy database. Proteome Discoverer search parameters included a mass range of 200- 5,000 Da, positive mode polarity, signal-to-noise ratio of 3, and a minimum peak count of one. SEQUEST search parameters were static modification of cysteine- carboxyamidomethylation, dynamic/variable modifications of methionine-oxidation, and phosphorylation of Ser, Thr, and Tyr residues. Other search parameters included two missed tryptic cleavage sites and precursor and fragment ion tolerances of 1.0 Da and 1000 ppm, respectively. The peptide filter of "charge state versus Xcorr" was enabled. Additional filters included peptide mass deviation of 50 ppm, maximum ΔCn of 0.01, and at least one peptide per protein. Phosphopeptides that passed these filters were further analyzed using phosphoRS algorithm of Proteome Discoverer (Taus et al. 2011). The FDR was fixed at 0.01.

**Proteomics Data Availability and Analysis**

The mass spectrometry proteomics data generated by experiment1 and experiment 2 were deposited to the ProteomeXchange Consortium (http://proteomecentral.proteomexchange.org) via the PRIDE partner repository (Vizcaino et al. 2013) with the dataset identifier PXD003676 (Reviewer account:

Username: reviewer09986@ebi.ac.uk Password: PvKeb5Sj). Experiment 3 data were deposited to ProteomeXchange Consortium with the dataset identifier PXD000000.

The proteins identified in both experiment 2 and experiment 3 with at least 2 unique peptides were used for GO analysis, KEGG pathway mapping, phylogenomic database search and cell wall protein database search. GO enrichment analysis was conducted at agriGO (http://bioinfo.cau.edu.cn/agriGO/analysis.php, v 1.2). For KEGG pathway search, we identified KO terms of all possible proteins identified with an online KO analysis tool at Rice Oligonucleotide Array Database (http://www.ricearray.org/analysis/overview.shtml). The identified KO terms are used for KO pathway search and mapping with KEGG mapper (http://www.genome.jp/kegg/mapper.html). We searched all possible protein models at specific rice GT and GH families, and at all transcription factor and kinases families in rice phylogenomic database (http://ricephylogenomics.ucdavis.edu/description.shtml). We also searched WallProtDB to find which previously identified extracellular proteins are present in our dataset (http://www.polebio.lrsv.ups-tlse.fr/WallProtDB/index.php).

**Metabolomics and Metabolite Identification**

Metabolites were extracted from frozen ground materials with a hot methanol-based method (Vanholme et al. 2012). Frozen samples were homogenized by Genogrinder and extracted with 1 mL methanol at 70ºC for 15 min. After 3 min centrifugation at 15000 rpm, 300 μL supernatant was transferred to a new tube and lyophilized for storage. Samples were dissolved with 400 μL of cyclohexane and water (1:1) before use.

For LC-MS/MS, two parallel experiments were conducted on positive electrospray ionization (ESI+) and negative electrospray ionization (ESI-) mode with all three biological replicates. Each sample was run in triplicate to ensure technical reproducibility. A pooled sample was made of equal amount of each sample and run in quintuplicate. For each run, 2 µL was injected for metabolite profiling using liquid chromatography with an ACQUITY HSS C18 column (2.1 x 150 mm 1.7 µm Particle) coupled with Waters SYNAPT G2-Si mass spectrometry. For UPLC, a gradient of two buffers was used: buffer A (99/1/0.1, water/acetonitrile/formic acid, pH 3.0, formic acid was not used for ESI- mode), buffer B (99/1/0.1 acetonitrile/water/ formic acid, pH 3.0); 95% A for 0.1 min decreased to 50 % A in 30 min (flow rate was 350 µL/min, column temperature was 40°C). For mass spectrometry, $MS^E$ was used for the acquisition mode, in which collision energy cycles between a low-energy state and a high-energy state, yielding fragment patterns for a large fractions of ions. Conditions were as follows: Mass range 50 to 1200 Da, scan speed 0.1 s per scan, source temperature 120°C, desolvation temperature 500°C, desolvation gas 1200 L/h, capillary 1.0 KV, cone voltage 20 V, source offset 80 V, cone gas 50 L/h, resolution 20000 (full width at half maximum), high energy collision ramp 30 to 50 eV.

Peaks were aligned, picked, normalized and identified using Progenesis QI for metabolomics analysis (v 2.1) with the settings stated below. Automatic peak picking was set to default sensitivity and the front signal before 2 min was removed. Every sample was normalized to an automatic determined reference sample based on all

compounds. We applied ChemSpider to search against a Plant metabolic pathway database (PlantCyc) and MetaScope to search against a custom rice metabolite database extracted from the Human Metabolome Database. For both methods, the identification parameters were: precursor mass tolerance 5 ppm, theoretical fragmentation was performed, fragment mass tolerance 5 ppm. Mass similarity, isotope similarity and fragmentation score were used to calculate the identification score for each peak (maximum 60). Fragmentation score was calculated based on the match with theoretical fragmentation (Wolf et al. 2010). Peaks with scores ≥50 were tentatively annotated with the highest score. The presences of *p*CA and apigenin in stem were verified by running standards and a pool of stem samples.

Statistical analyses, including principal component analysis, analysis of variance (ANOVA), and k-means clustering were performed in R (v 3.12). For all analyses, we only used metabolite ions that are: (1) Consistently present or absent in the technical replicates of all samples (with an average consistency score of at least 0.4 in all sample). $Consistency\ score = |\frac{s}{n} - 0.5|$ , where $s$ is the number of presence in $n$ technical replicates. For example, if the metabolite is present or absent in 4 out of 5 technical replicates of the pool sample, it has a consistency score of 0.4. Technical replicates were average before analysis. For k-means clustering, we first performed k-means clustering with k = 2-20 and compared the silhouette plot and average silhouette width of results. We used k = 5 for clustering since it gave high silhouette coefficients in most clusters and had fewer outliers. Sample ES-3 and PMS-1 were slightly out of stage when collected, and therefore excluded from the ANOVA and clustering analysis.

**Figures and Tables**

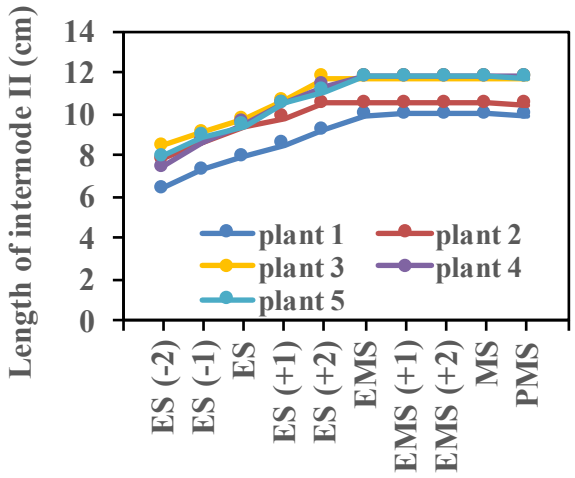



**Figure 3-1** The internode II elongates until the panicle fully emerges from the leaf sheath. Internode lengths were measured across four developmental stages: ES, elongating stem, the panicle is about to emerge from the flag leaf sheath; EMS, early mature stem, panicle fully emerged from the flag leaf sheath; MS, mature stem, green husks and milky grains; PMS, post mature stem, yellow husks and hard grains. Negative numbers and positive numbers in brackets indicate days before and after a stage, respectively.

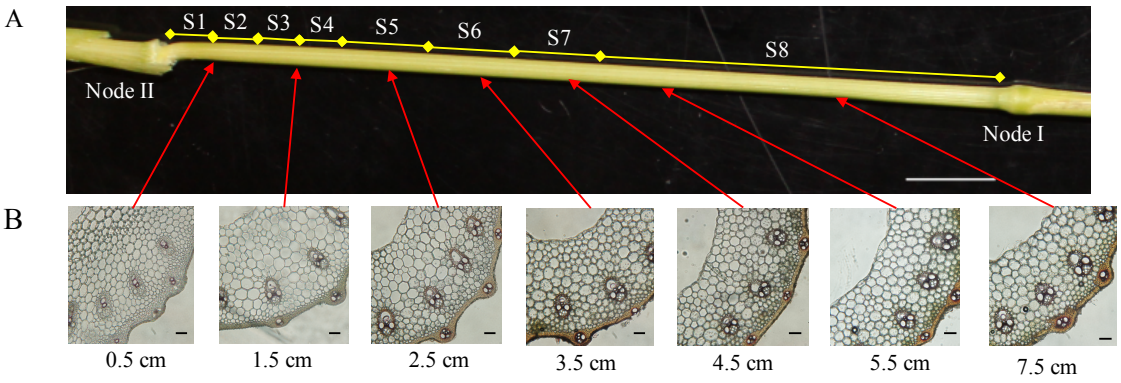



**Figure 3-2.** Internode II samples used for cell wall analysis. (A) Internode II of elongating stem at boosting stage when the panicle is just about to emerge. Scale bar is 1 cm (B) Phloroglucinol stain of cross-sections taken from the bottom (left) of the elongating internode to the top (right) indicate the gradual accumulation of lignin. The labels below each photo indicate the distance from node II. Scale bar is 100 µm.

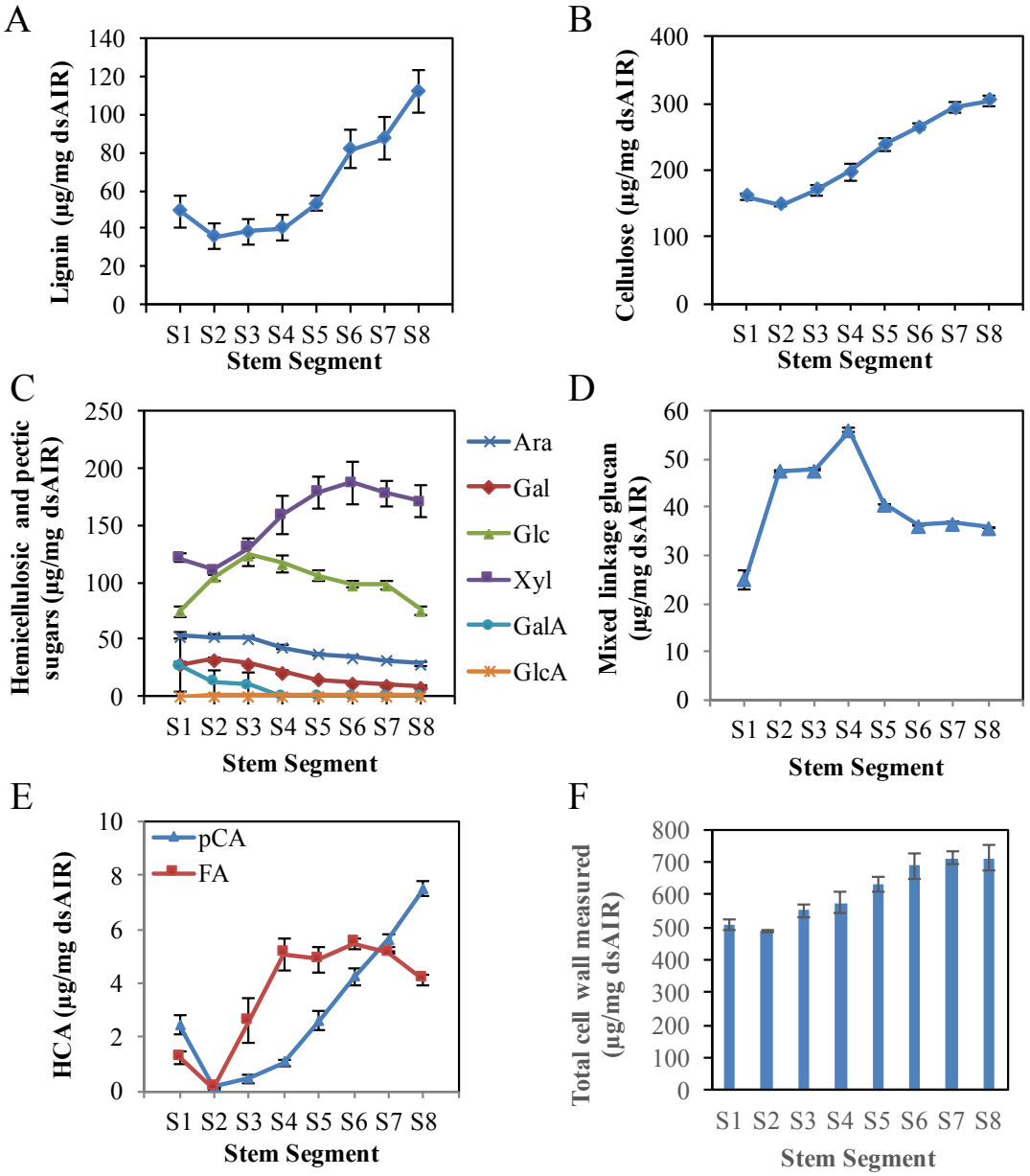



**Figure 3-3.** Cell wall composition changes along the rice elongating internode. The internode was divided into eight uneven segments, from the youngest segment (S1) to the oldest (S8) (See Methods). Cell wall components are presented as percent weight of de-starched alcohol insoluble residue (dsAIR). Error bars indicate standard errors of three biological replicates for all measures except lignin, which used four replicates. **(A)**

Acetyl bromide soluble lignin. **(B)** Cellulose measured by anthrone assay. **(C)** Hemicellulosic and pectic sugars. Ara, arabinose; Gal, galactose; Glc, glucose; Xyl, xylose; GalA, galacturonic acid; GlcA, glucuronic acid. **(D)** Mixed linkage glucan measured by β-glucan assay kit **(E)** Cell wall-associated hydroxycinnamic esters (HCAs) including ferulic acid (FA) and *p*-coumaric acid (*p*CA) (**F**) The sum of the mass of cell wall components measured except mixed linkage glucan.

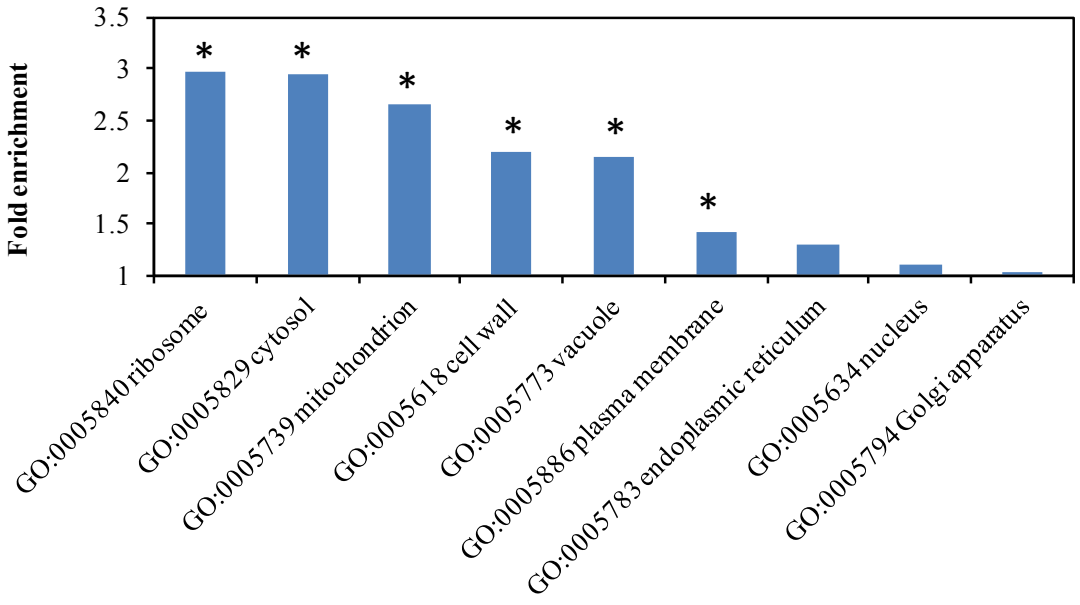



**Figure 3-4**. Representation of cellular component Gene Ontology (GO) terms among rice elongating internode proteins. Star indicate over-representation with a false discovery rate <0.01 by hypergeometric test. Fold enrichment is the ratio of identified proteins to the expected number of proteins for each GO terms.

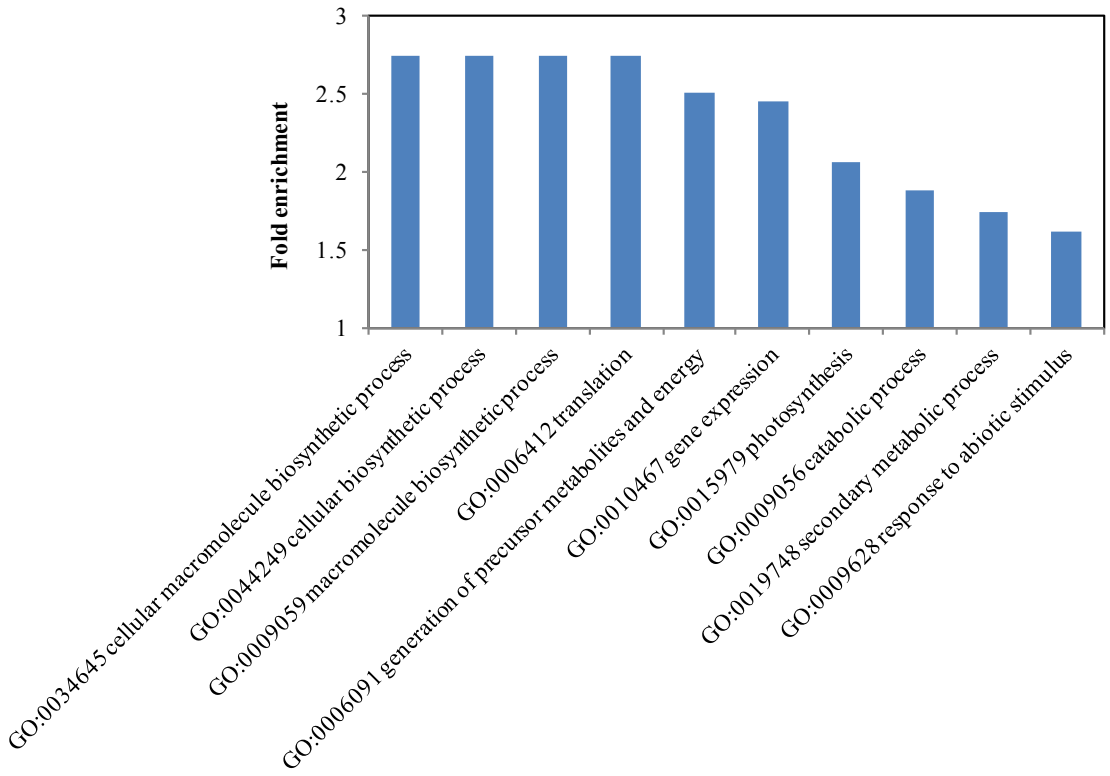



**Figure 3-5.** Top 10 over-representation of biological process Gene Ontology (GO) terms among rice elongating internode proteins. All terms are over-represented with a false discovery rate <0.01 by hypergeometric test. Fold enrichment is the ratio of identified proteins to the expected number of proteins for each GO terms.

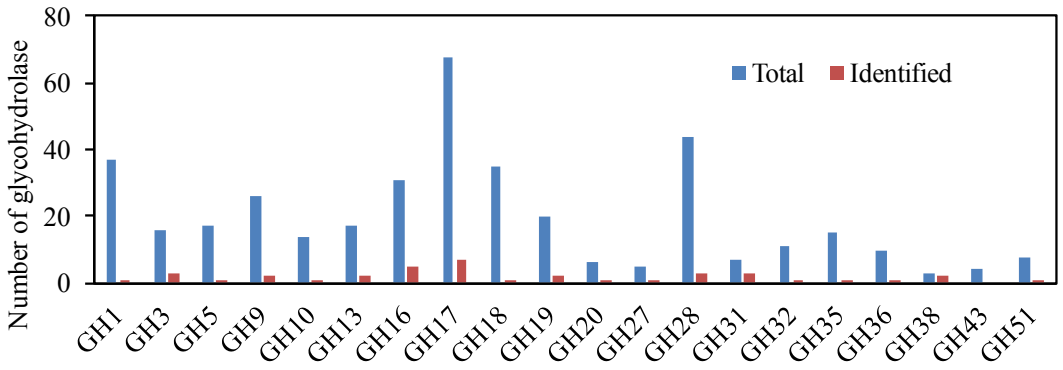



**Figure 3-6.** The representation of identified proteins in glycoside hydrolases (GH) families that may modify or degrade grass cell walls. Red bars indicate the number of GH proteins identified in this study. Blue bars indicate total number of GH proteins in rice GH database. GH families that modify pant cell wall components or localize to cell wall or expand in monocots are included.

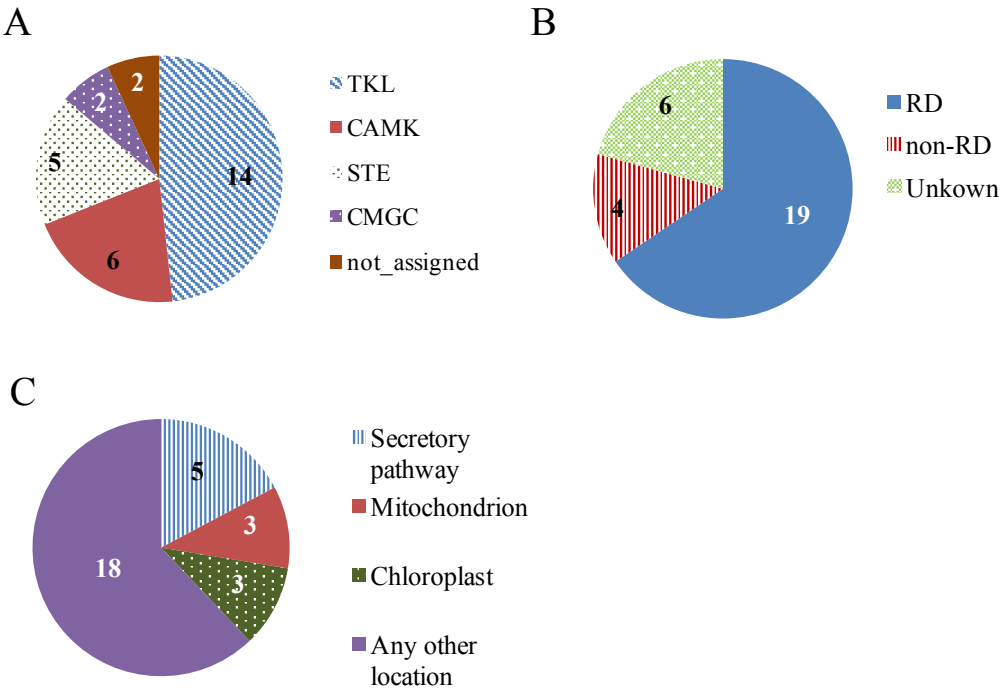



**Figure 3-7.** Protein kinases in the rice elongating internode. **(A)** Kinases by phylogenetic groups. The TKL (Tyr kinase-like) group includes mixed lineage kinases, transforming growth factor-β receptor kinases, and Raf kinases. The CAMK group consists of calcium/calmodulin-dependent protein kinases. The STE group includes homologs of yeast sterile 7, sterile 11, and sterile 20 kinases. The CMGC group includes CDK, MAPK, GSK3, and CLK kinases. The AGC group includes PKA, PKG, and PKC kinases. **(B)** Kinase categories based on a conserved kinase RD motif. Kinases containing this motif are RD kinases while non-RD kinases lack the motif. **(C)** Predicted subcellular localization of kinases.

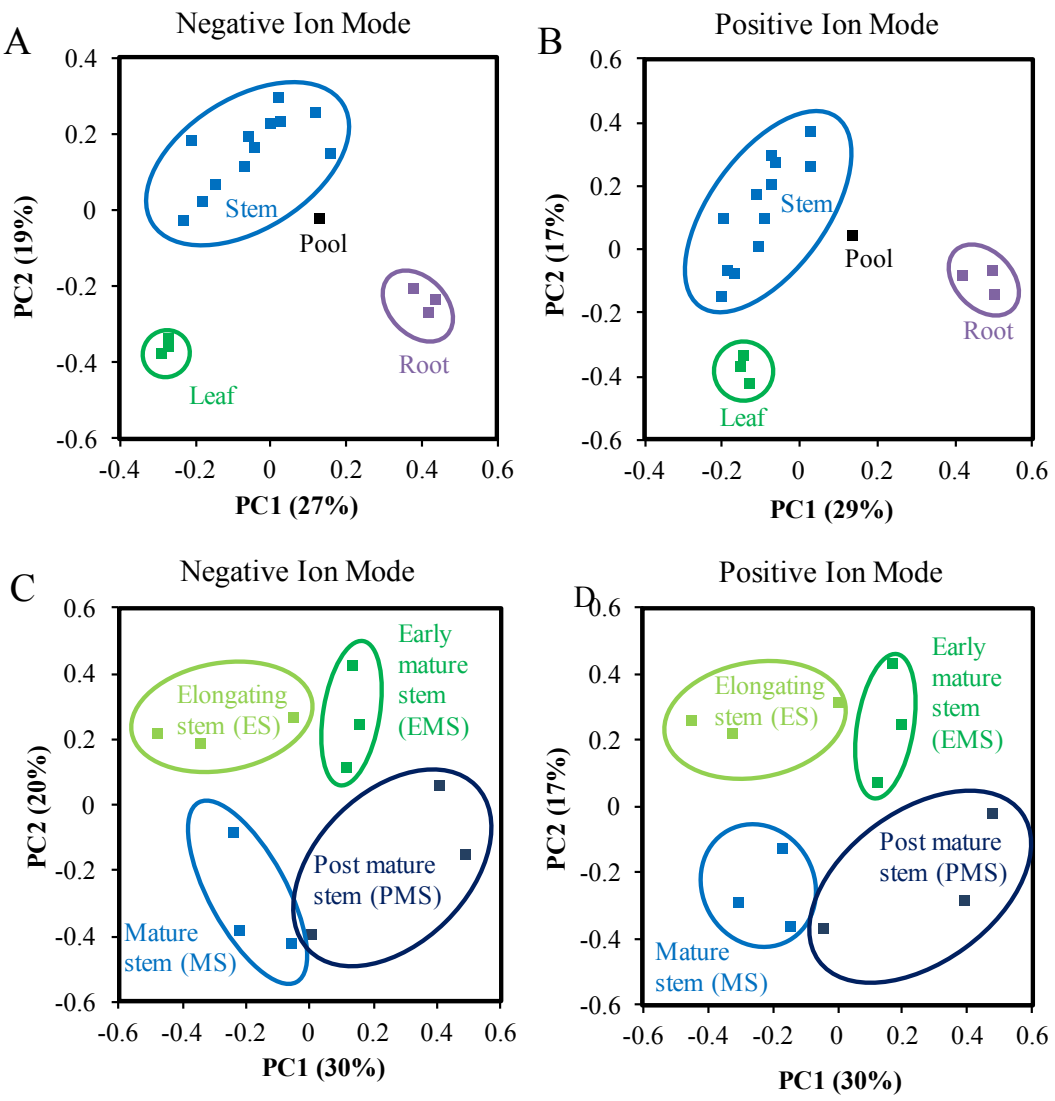



**Figure 3-8.** Methanol extracted metabolite profiles differ among rice organs and internode II of different stem development stages. **(A &B)** Principal component analysis (PCA) of samples from different organs. Leaf samples were labeled green, root are purple, stem are blue. The pool sample was labeled black. **(A)** Negative ion mode. **(B)** Positive ion mode. **(C &D)** PCA of stem samples of different developmental stages. Light green, green, blue, and dark blue represent elongating stem internode (ES), early

mature stem internode (EMS), mature stage stem internode (MS), and post mature stem internode (PMS), respectively. **(C)** Negative ion mode. **(D)** Positive ion mode.

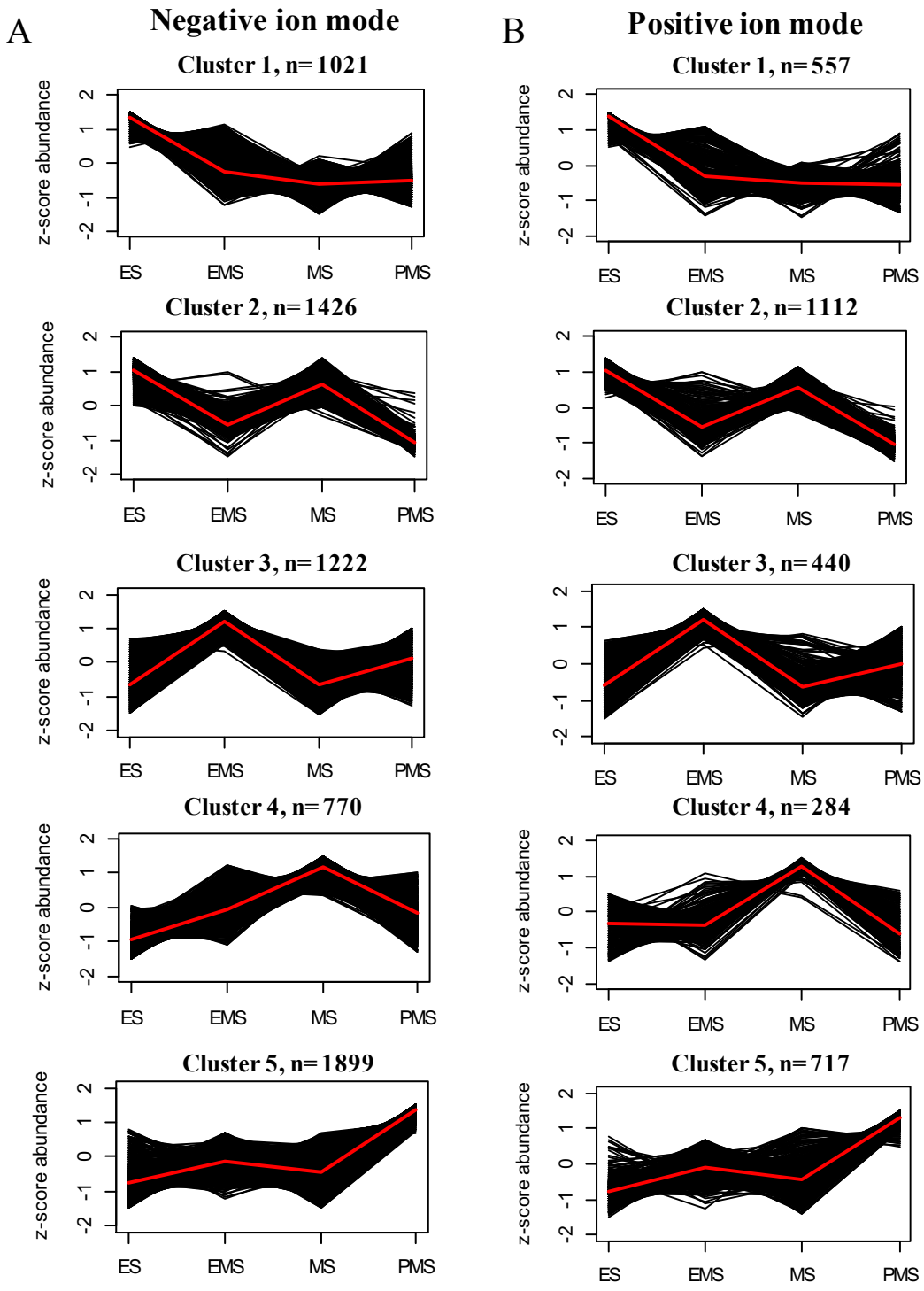



**Figure 3-9**. K-means clusters of metabolites with significant change during stem internode maturation. The four stages are: elongating stem (ES), early mature stem

(EMS), mature stem (MS), post mature stem (PMS). Metabolite ions with a *q*-value < 0.01 by ANOVA test were used for clustering analysis. The total number of clusters was determined by silhouette plot. Metabolites change in similar pattern during the four stages were grouped into five cluster. Black lines show the relative abundance of all metabolite ions in each cluster and the red line shows the average of them. The total number of metabolite ions in each cluster is indicated by n **(A)** Negative ion mode **(B)** Positive ion mode

Table 3-1. Samples and methods for the three proteomics experiments in this study.

| | Experiment 1 | Experiment 2 | Experiment 3 |
|---|---|---|---|
| Sample | Elongating internode rep 1 [a] | Same as experiment 1 [b] | Elongating internode rep 2 [a] |
| Technical replicates | 3 | Same as experiment 1 | 4-plex isobaric labeling |
| Protein extraction | Tris buffered phenol | Same as experiment 1 | $NH_4Ac$ buffered MeOH/β-mercaptoethanol |
| Trypsin digestion | In-gel digestion | Same as experiment 1 | Buffered protein suspension |
| Protein fractionation prior to LC-MS/MS | Gel electrophoresis followed by gel slicing | Same as experiment 1 | Off line fraction collection usi high-pH C18 reversed phase L |
| Liquid chromatography | C18 reversed phase | C18 reversed phase | C18 reversed phase |
| Mass spectrometer | LTQ Orbitrap | SYNAPT G2-S*i* HDMS | Q Exactive Plus |
| Data acquisition mode | Data dependent acquisition | Data independent acquisition | Data dependent acquisition |
| Peptide search | Proteome Discoverer v1.4, SEQUEST algorithm | Progenesis QIP v2.0 | MS-GF+ v2017.01.13 |
| Trypsin mis-cleavage | 2 | 1 | Full or partial tryptic ends including missed cleavages |
| Peptide mass tolerance (ppm) | 25 | 20 | 10 |
| Peptides per protein | ≧1 | ≧2 | ≧2 |
| Peptide false discovery rate | <5.1% | <4% | ≦1% |

[a] Internode II of ES is defined as the elongating internode. Elongating internode rep 1 and rep2 were biologically matched samples.

[b] Experiment 1 and experiment 2 used

**Table 3-2. Rice phenylpropanoid enzymes observed in the elongating internode.**

| Locus ID [a] | Gene family [b] | Peptide count [c] | Unique peptides [d] | Function or putative function | Citation |
|---|---|---|---|---|---|
| LOC_Os02g41630 | PTAL | 83 | 50 | Putative phenylalanine /tyrosine ammonia-lyase | |
| LOC_Os05g25640 | C4H | 3 | 2 | Putative cinnamic acid 4-hydroxylase | |
| LOC_Os02g08100 | 4CL | 33 | 21 | 4CL3, 4-coumarate:CoA ligase | (Gui et al. 2011) |
| LOC_Os06g44620 | 4CL | 24 | 10 | 4CL4, putative 4-coumarate:CoA ligase | (Gui et al. 2011) |
| LOC_Os02g39850 | HCT | 10 | 8 | HCT2, hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase | (Kim et al. 2012) |
| LOC_Os04g42250 | HCT | 6 | 4 | Putative hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase | |
| LOC_Os06g06980 | CCoAO | 14 | 14 | Putative caffeoyl CoA 3- | |

| | | | | | |
|---|---|---|---|---|---|
| | MT | | | O-methyltransferase | |
| LOC_Os05g41440 | C3H | 3 | 2 | Putative *p*-coumaroyl shikimate 3′-hydroxylase | |
| LOC_Os08g34280 | CCR | 8 | 5 | CCR1, cinnamoyl CoA reductase | (Kawasaki et al. 2006) |
| LOC_Os02g09490 | CAD | 25 | 25 | CAD2, cinnamyl-alcohol dehydrogenase | (Zhang et al. 2006) |
| LOC_Os09g23540 | CAD | 20 | 4 | Putative cinnamyl-alcohol dehydrogenase | |
| LOC_Os08g02110 | POX | 9 | 8 | Putative class III plant peroxidase [e] | |
| LOC_Os02g58720 | POX | 8 | 7 | Putative class III plant peroxidase [e] | |
| LOC_Os06g46799 | POX | 5 | 4 | Putative class III plant peroxidase [e] | |
| LOC_Os09g29490 | POX | 5 | 4 | Putative class III plant peroxidase [e] | |
| LOC_Os01g36240 | POX | 3 | 2 | Putative class III plant peroxidase [e] | |
| LOC_Os01g40860 | REF | 14 | 8 | Putative sinapaldehyde dehydrogenase | |

[a] Rice Genome Annotation Project (Version 7) (Ouyang et al. 2007).

[b] Gene families involved in phenylpropanoid biosynthesis. PTAL, phenylalanine/tyrosine ammonia-lyase; C4H, cinnamic acid 4-hydroxylase; 4CL, 4-coumarate:CoA ligase; HCT, hydroxycinnamoyl-Coenzyme A shikimate/quinate hydroxycinnamoyltransferase; CCoAOMT, caffeoyl CoA 3-O-methyltransferase; C3H, *p*-coumaroyl shikimate 3′-hydroxylase; CCR, cinnamoyl-CoA reductase; CAD, cinnamyl alcohol dehydrogenase; POX, peroxidase; REF, reduced epidermal fluorescence.

[c] Total number of peptides identified for a protein including those that match with other proteins.

[d] The number of unique peptides private to a protein.

[e] Class III peroxidases belong to one of the more than 60 classes of peroxidases as described in the PeroxiBase (updated May 2015, http://peroxibase.toulouse.inra.fr/). Class I-III peroxidases are found only in plants,

**Table 3-3. Rice cell wall-related acyltransferases and glycosyltransferases observed in the elongating internode.**

| Locus ID [a] | Protein [b] | Peptide count [c] | Unique peptide [d] | Function or putative function | Citation |
|---|---|---|---|---|---|
| LOC_Os01g09010 | OsAT9 | 27 | 22 | Putative feruloyl coenzyme A transferase involved in glucuronoarabinoxylan. | (Lin et al. 2016) |
| LOC_Os01g42880 | OsAT1 | 26 | 22 | Involved in the feruloylation of glucuronoarabinoxylan in Brachypodium. | (Buanafina et al. 2016) |
| LOC_Os01g42870 | OsAT2 | 11 | 10 | Putative feruloyl coenzyme A transferase | (Molinari et al. 2013) |
| LOC_Os05g04584 | OsAT3 | 11 | 9 | Involved in the *p*-coumaroylation of lignins | (Sibout et al. 2016) |
| LOC_Os06g39390 | OsAT10 | 9 | 9 | *p*-coumaroyl coenzyme A transferase involved in glucuronoarabinoxylan modification | (Bartley et al. 2013) |

| | | | | | |
|---|---|---|---|---|---|
| LOC_Os01g18744 | PMT /OsAT4 | 5 | 5 | *p*-coumarate monolignol transferase contributes to p-coumaroylation of lignins | (Petrik et al. 2014; Withers et al. 2012) |
| LOC_Os05g08370 | GT2-CESA1 | 5 | 5 | Cellulose synthase | |
| LOC_Os03g60939 | GT2 | 2 | 2 | Putative glycosyltransferase | |
| LOC_Os02g22380 | GT61-XAX1 | 15 | 14 | Glycosyltransferase involved in xylan side chain synthesis. | (Chiniquy et al. 2012) |
| LOC_Os06g27560 | GT61-IV-1 | 15 | 14 | Putative glycosyltransferase | |
| LOC_Os01g02900 | GT61-IV-2 | 2 | 2 | Putative glycosyltransferase | |
| LOC_Os02g22190 | GT61-III-1 | 4 | 2 | Putative glycosyltransferase, high expression in root and stem at transcript level | |

[b] Acyltransferases (AT) and glycosyltransferases (GTs) involved in polysaccharide biosynthesis. PMT, *p*-coumaroyl-CoA:monolignol transferase; CESA, cellulose synthase A; IRX, irregular xylem; XAX, xylosyl arabinosyl substitution of xylan. GT61 subclade III and VI were defined in Chiniquy, 2012.

[c] Total number of peptides identified for a protein including those that match with other proteins.

[d] The number of unique peptides private to a protein.

**Table 3-4.** Metabolites enriched at different stages of rice stem internode development

| | | ESI- [c] | | ESI+ [c] | |
|---|---|---|---|---|---|
| K-means cluster [a] | Enriched sample(s) [b] | Metabolite ions [d] | Representative metabolite [e] | Metabolite ion [d] | Representatives metabolite [e] |
| Cluster 1 | Elongating stem (ES) | 1021 | coniferyl aldehyde | 667 | methyl cinnamate |
| Cluster 2 | Elongating stem (ES) and mature stem (MS) | 1426 | NI | 1112 | 6,7-dihydroxycoumarin, phenylacetaldehyde |
| Cluster 3 | Early mature stem (EMS) | 1222 | NI | 440 | demethoxycurcumin, xanthotoxol |
| Cluster 4 | Mature stem (MS) | 770 | NI | 284 | benzaldehyde, olivetol |
| Cluster 5 | Post mature stem (PMS) | 1899 | isoscoparin 2-(6-(E)-p-coumaroylglucoside), isoscoparin 2-(6-(E)-ferulylglucoside) | 717 | tricin 7-glucoside |

[a] Clusters of metabolites were determined based on abundance within the dataset.

[b] Sample or samples in which metabolites in the cluster are most abundant.

[c] Charge of the electrospray used to ionize metabolites prior to mass spectrometry. ESI-, negative electrospray ionization mode; ESI+, positive electrospray ionization mode.

[d] The number of metabolite ions in the cluster. The retention time and mass-to-charge ratio of detected ions can be found in Table 3-S10 and S11. NI indicates no confident identifications.

[e] Examples of tentatively annotated metabolites. Other annotated metabolites can be found in Table 3-S10 and S11.

***Supplementary data***

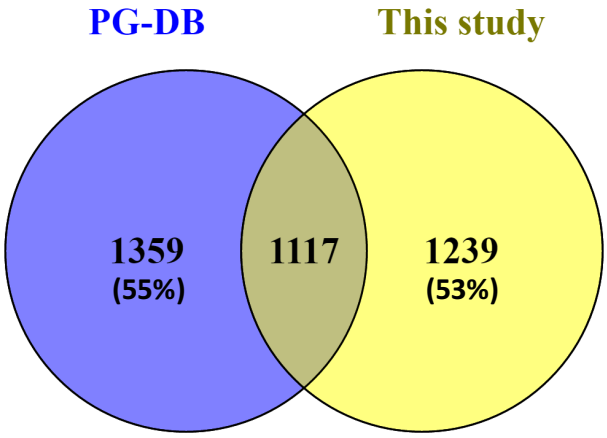



**Figure 3-S1** Proteins identified in this study partially overlap with the proteins in Rice Proteogenomics Database (Oryza PG-DB). Proteins with at least 2 peptides from Oryza PG-DB were compared to proteins identified by at least 2 unique peptides in this study. Percentages indicate the percentage of proteins only presented in each dataset.

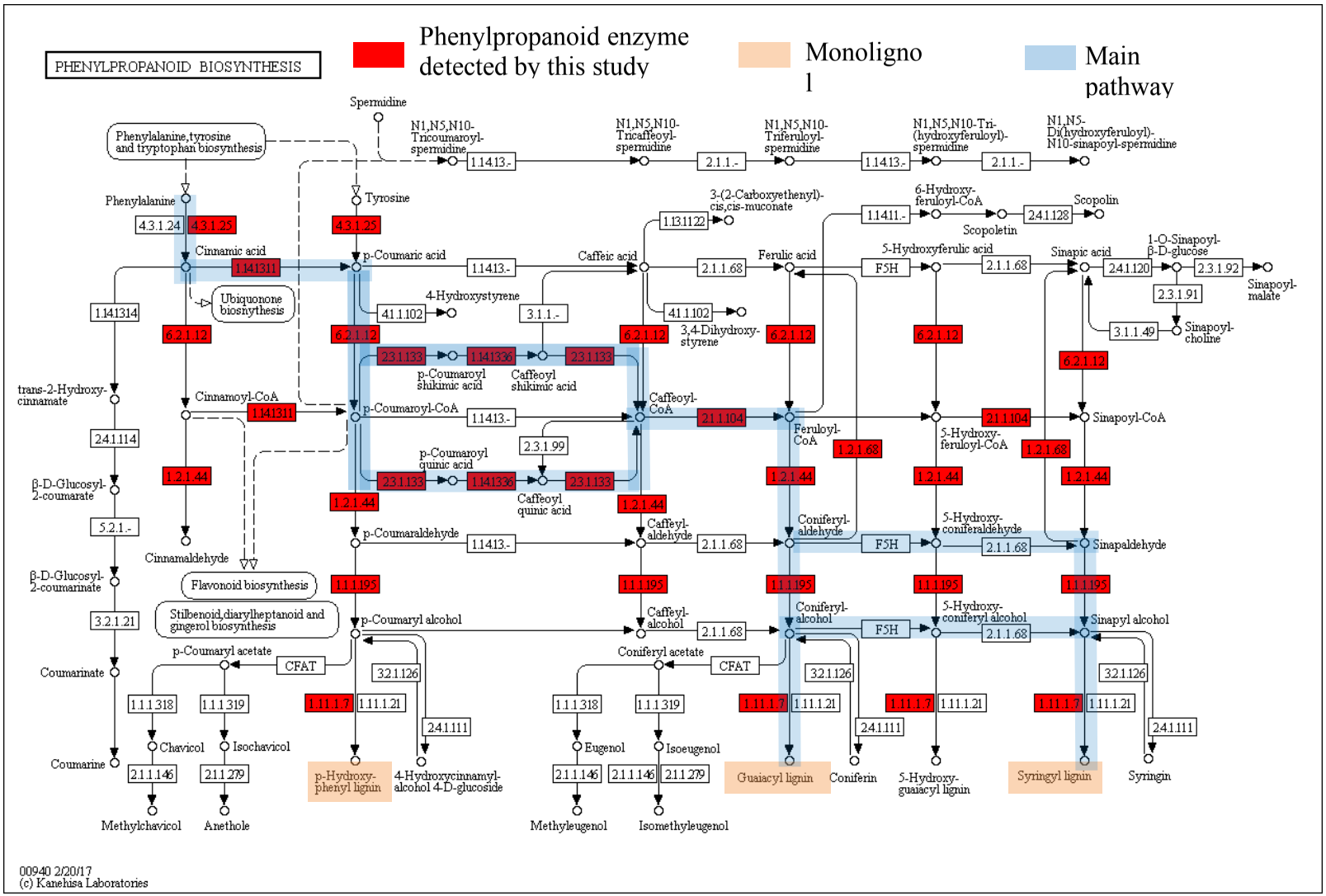



**Figure 3-S2**. Most phenylpropanoid enzymes were identified in rice elongating internode II. Identified proteins were mapped to the KEGG phenylpropanoid pathway (osa00940). Red boxes are phenylpropanoid synthesis enzymes identified by this study. Key lignin synthesis reactions were labeled by blue shading. Orange shadings indicate three common lignin

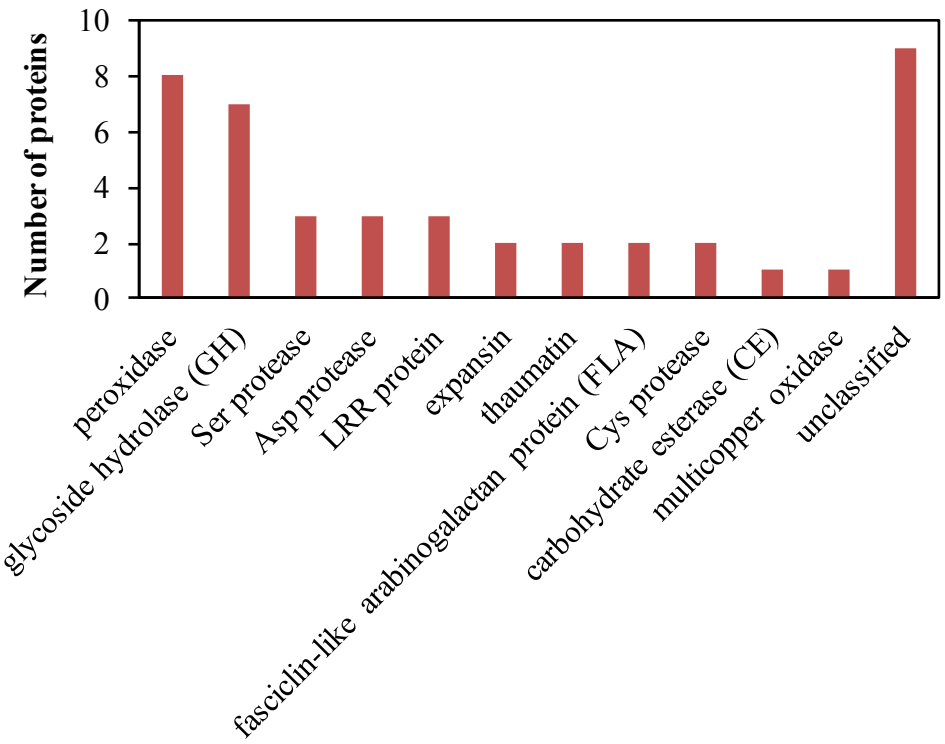



**Figure 3-S3.** Presence of previously reported extracellular proteins in the rice elongating internode. Proteins identified by this study were searched against WallProtDB database for extracellular proteins that have been reported by cell wall proteomics.

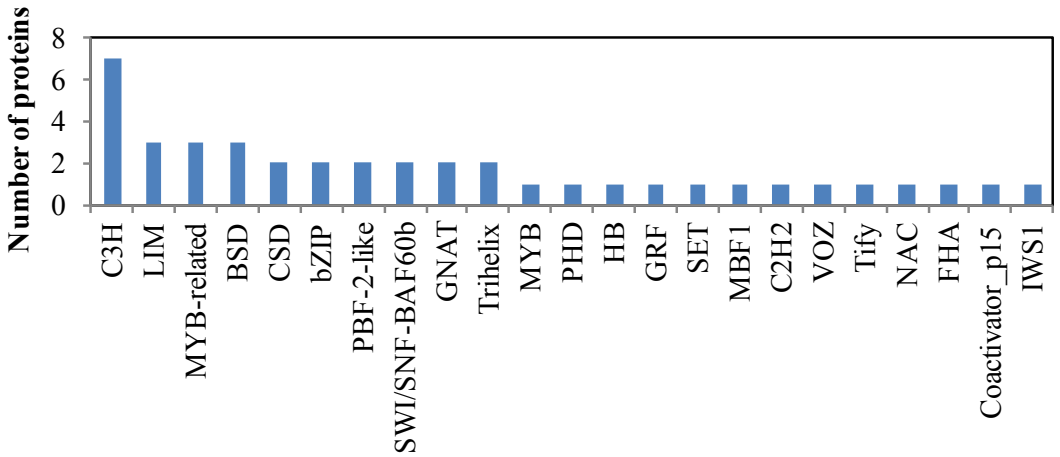



**Figure 3-S4.** Transcription factors identified in the rice elongating internode. Identified proteins were searched against a rice TF database for transcription factors.

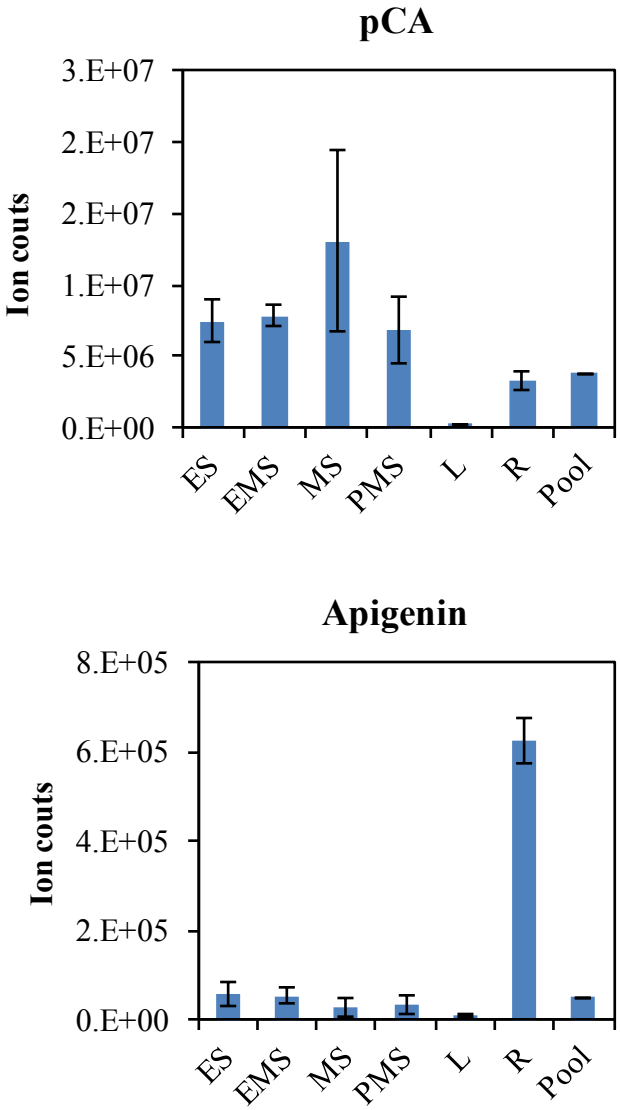



**Figure 3-S5.** The abundance of *p*-coumaric acid (*p*CA) and apigenin differ among organs. Error bars are standard deviation among three biological replicates.

The following supplementary table will be available at Frontiers in Plant Science

**Table 3-S1.** Cell wall composition in internode segments.

**Table 3-S2.** Protein identified with at least 2 unique peptides by both experiment 2 and experiment 3.

**Table 3-S3.** Biological process, molecular function, and cellular component GO terms enriched in the rice elongating internode.

**Table 3-S4.** KEGG ontology analysis of identified proteins in the elongating internode.

**Table 3-S5.** Identified glycoside hydrolases (GHs) from the rice GH database.

**Table 3-S6.** Presence of previously identified proteins by cell wall proteomics..

**Table 3-S7.** Characteristics of rice kinases identified in the elongating internode.

**Table 3-S8.** Transcription factors detected in the elongating internode.

**Table 3-S9.** Phosphopeptides identified in the rice elongating internode.

**Table 3-S10.** Relative abundance of negative ions in four stages of rice stem internodes.

**Table 3-S11.** Relative abundance of positive ions in four stages of rice stem internodes.

# Chapter 4: The Effect of Genetic Modification of The Lignin Biosynthesis Pathway on Low Temperature Pyrolysis Product Yields

**Authors:** Christopher Waters, Fan Lin, Laura Bartley, Richard Mallinson, Lance Lobban

**Publication Status**: This chapter is extracted from a manuscript prepared for Green Chemistry

**Author contributions:** C.W. designed this study and conducted the thermal conversion experiment. F.L. prepared biomass samples, measured HCA content in COMT mutant, and conducted the alkali treatment. C.W and F.L. drafted the manuscript.

**Abstract**

Thermochemical conversion can rapidly and efficiently convert biomass to liquid fuel precursors, which can be upgraded to transportation fuels by catalytic reactions. Since convention catalytic strategies can not handle all classes of thermal products as a mixture, separation of thermal products is necessary. Thermal fractionation is an approach to separate products by applying successively increasing series of temperatures. However, the broad decomposition temperature range of lignin hinders the segregation of lignin-derived products from polysaccharide-derived products. Therefore, reducing lignin decomposition at low temperature (290°C), at which polysaccharides decompose, will enhance the efficiency of thermal fractionation. In this study, we examined a down-regulation mutant of Caffeic acid O-methyltransferase (COMT) that synthesize the precursor of lignin S subunits. We hypothesized the reduction of S subunits in the COMT mutant as reflected by a decreased ratio of S subunit to G subunit (S/G ratio) will reduce low energy bonds in lignin and therefore increase the thermal stability of lignin. We measured the yield of phenolic products, which are mostly from lignin and hydrocinnamic acid (HCAs), at low (350°C) and high conversion temperatures (500°C). The mutant biomass has decreased yield of phenolic products at low temperature and no significant change of total phenolic product yield comparing to the wild-type. Though HCAs, which are less thermally stable, also decrease in the mutant, a reduction of phenolic product yield at low temperature is still observed after HCAs being removed from biomass by an alkali-treatment. We therefore conclude that reducing lignin S subunits can reduce lignin decomposition at low temperature and facilitates better segregation of lignin derived-thermal products from

polysaccharide-derived products. This study demonstrated genetic manipulation on lignin structure can help improving the efficiency of thermal fractionation.

## Introduction

Thermochemical conversion of biomass, and especially biomass fast pyrolysis, has been projected to be the most efficient method producing drop-in renewable fuels in the mid- to long-term. As compared to enzymatic conversion technologies, the rates of reaction afforded by the very high temperatures involved and the elimination of the high fresh water requirements drive these efficiency gains (Verma et al. 2012). Despite this, the significant downstream processing required to convert biomass pyrolysis products into refinery-compatible product streams remains a barrier to industry adoption of these technologies (Tanger et al. 2013). Due to the myriad issues involved with the liquid handling of condensed biomass pyrolysis products (hereafter referred to as 'bio-oil'), some form of catalytic treatment is necessary, primarily to remove oxygen which is responsible for many of the undesirable properties of bio-oil including its reactivity, acidity, and low heating values.

A variety of catalytic strategies are used to treat biomass pyrolysis products; some target the uncondensed vapor products (in-situ or ex-situ catalytic fast pyrolysis), while others are designed for liquid-phase bio-oil. Earlier work in pyrolysis product valorization involved investigation into the suitability of familiar petrochemical catalytic chemistries (most commonly zeolite cracking or hydrodeoxygenation) (Bridgwater 1994; Carlson et al. 2009). The resulting product streams are usually refinery compatible but suffer from poor efficiencies or economics. For example, hydrodeoxygenation could be very suitable for conversion of lignin derived phenolics, typically with 6 to 9 C-C bonds and high C:O ratios, but the high amounts of lighter,

C1-C4 polysaccharide-derived oxygenated species present lead to poor overall liquid carbon retention and prohibitive hydrogen consumption. Similarly, zeolites can effectively convert lighter oxygenates into BTX streams but rapidly deactivate due to internal and external coking reactions, causing much of the carbon in biomass to be lost as CO2 during catalyst regeneration (Carlson et al. 2008). To improve catalyst performance, some separation of the pyrolysis products must be achieved prior to reaction.

One strategy proposed in the literature for segregation of biomass pyrolysis products is the use of a successively increasing series of temperatures to fractionate the biomass into product streams of approximate composition of those produced by individual biopolymers (Herron et al. 2017). This approach, using thermal segregation only, is attractive when compared with other options such as hybrid thermal-solvolysis due to low requirements for process chemicals or costly solid-liquid operations.

However, a challenge for this approach is the broad temperature decomposition range of lignin. While most hemicelluloses and cellulose typically thermally decompose at separate temperature ranges, lignin is known to begin to decompose at temperature ranges which span those of both polysaccharides, causing the characteristic phenolic products to be released into multiple thermal streams (Antal and Varhegyi 1995; Varhegyi et al. 1997). This is especially troublesome for the lowest-temperature regimes targeted at decomposing hemicellulose. Hemicellulose degradation yields most of the acetic acid produced in pyrolysis, and acetic acid is known to catalyze

oligomerization reactions with these heavy phenolics (Tanger et al. 2013). In addition to this storage & handling concern, hemicellulose products are in general well suited for upgrading via zeolites or using aldol condensation/ketonization catalysts, both of which are deactivated by lignin-derived phenolics (Cheng et al. 2016).

While the use of a low-lignin feedstock could help to minimize the selectivity to lignin-derived phenolics in a first stage product stream, high amounts of lignin-derived phenolics are desirable from a whole-process viewpoint as the high C:O ratio and abundant C-C bonds could lead to relatively low-input fuel range molecules. However, if the thermal stability of the lignin could be altered to suppress decomposition at lower temperatures, especially without reducing the total lignin amount, both of these desirable outcomes could be achieved.

The broad range of the thermal stability of lignin arises from the variety of different types of linkages present in the polymer. Lignin is a complex, cross-linked polymer comprising primarily three monomer (H, G, and S) units cross-coupled via oxidative reactions during incorporation in the cell wall (Boerjan et al. 2003). Figure 4-S1 shows an overview of these three lignin monomers and their resulting structures in the polymer. The most frequently occurring polymer unit is the β–O–4 structure (A), accounting for more than half of the inter-unit linkages; other commonly occurring units are β–5 (B), β–β (C), 5–5 (D), 5–O–4 (E), and β–1 (F) (other units are various special cases, and are included by the author for completeness). Increased production of G monomers over S monomers (i.e., a lower S/G ratio) leads to an increased prevalence of

β–5, 5-5, and 5-O-4 linkages in the lignin polymer over β–O–4, β–β, and β–1 linkages due to the increased availability of the 5 position as a reaction site (Boerjan et al. 2003). Table 4-1 summarizes the results from Density functional theory (DFT) calculations of the enthalpy of dissociation for these six bonds. These suggest that, in general, due to the increased prevalence of high-stability linkages with low S/G ratios, we should expect the initiation of the radical depolymerization reactions of the lignin biopolymer to be slower at lower temperatures for low S/G ratio lignins. This would result in lower selectivity to lignin decomposition products at lower temperatures.

As lignin synthesis is a biological process, it includes multiple steps of reaction. The production of all three monomers is determined by several enzymes in the phenylpropanoid synthesis pathway (Boerjan et al. 2003). One of them, Caffeic acid O-methyltransferase (COMT) is responsible for the methylation on 5-OH group to produce sinapaldehyde and sinapyl alcohol, the precursors of S lignin monomer. The knockdown mutants of COMT in Arabidopsis have lower S/G ratio as a result of the reduction of S monomer and relative elevation of guaiacyl and 5-hydroxyl guaiacyl monomer. In this mutant, NMR data indicate inter-unit linkages such the β-aryl ether (β-O-4) (A) and resinol (β-β) (B) decreased but benzodioxane (β-O-4) increased (Vanholme et al. 2010b). The COMT mutants in bioenergy crops such as switchgrass and sorghum also show lower S/G ratio phenotype, which indicates COMT plays a similar role across dicot plants and monocot plants.

Switchgrass is an excellent bioenergy crop because it has high biomass production, low requirement for water and nutritional input, good tolerance to drought and high photosynthesis efficiency. Two-year field study indicates the knockdown of COMT in switchgrass show 20-30% reduction in lignin S/G ratio and 10% reduction of total lignin in senesced plants (Baxter et al. 2014). Polysaccharides in the mutants are mostly unchanged expect a 6-8% elevation of xylans. At the secondary year of growing in fields, the COMT mutants did not show reduction in dry biomass yield, tiller height, plant diameter, tiller number and disease resistance, which indicate plant growth and productivity is not sacrificed for the change in lignin composition. These characteristics of COMT mutant make it suitable for improving biofuel production.

In this study, we compare the low-temperature and high-temperature thermochemical products of the genetically modified switchgrass COMT mutants to those of the unmodified wild-type. We hypothesize that this modification will result in lower phenolic product yield during low temperature pyrolysis for the down-regulated mutants as compared with their unmodified wild-type.

**Results and Discussion**

Since HCAs contribute to the yield of phenolics products and are decreased in a sorghum COMT mutant (Palmer et al. 2008), we first checked the HCA content in the switchgrass COMT mutant. A significant decrease of both FA and *p*CA are observed. Though FA only has a subtle decrease in the COMT mutant, *p*CA decreases more than 30% (Figure 4-1). The decrease of HCAs is likely to be a side effect of COMT down-regulation, which may impact the phenylpropanoid pathway that synthesizes both lignin precursors and the putative precursors of cell wall incorporated HCAs. A previous transcriptomics and metabolomics study indicates the phenylpropanoid pathway are suppressed in the Arabidopsis COMT mutant and other S-lignin mutants (Vanholme et al. 2012). It has been proposed there is a feedback system to adjust the phenylpropanoid pathway in response to the modifications on lignin though the regulatory mechanisms for this system is mostly unknown (Vanholme et al. 2012).

We did thermal conversion experiment on the fine-ground biomass of two COMT mutant lines named as COMT-2 and COMT-3. To separate the factors that can contribute to the yield of phenolic products, we included biomass samples treated by NaOH and water. The water treatment will remove soluble phenolic metabolites in biomass and the NaOH treatment will remove both soluble phenolic metabolites and the HCAs incorporated to cell walls. A two-step thermal conversion was conducted on the treated and untreated biomass with the low temperature regime at 290°C and the high temperature regime at 500°. Thermal product yields at both steps were measured by a GC/MS-FID system. Since the lignin in COMT mutant is 10% less than the wild type

(Baxter et al. 2014), the yield of phenolic products was normalized by the weight of lignin.

For untreated biomass samples, the mutant biomass yield 30% less phenolic products than the wild type at low temperature (290℃) (Figure 4-2A). The phenolic product yield at high temperature (500℃) is only significantly decreased in the COMT-2 mutant line but not in the COMT-3 mutant line (Figure 4-2 B). The changes of total phenolic product yield in the mutant (Figure 4-2 C) was similar to that of the high temperature regime since more than 80% of the total phenolic products were produced at the high temperature step.

Comparing to untreated samples, the water treated samples yield 50% less phenolic products and the NaOH treated samples yield 80% less phenolic products at low temperature (Figure 4-2A). However, the phenolic product yield at high temperature remains the same after both treatments. This indicates HCAs and soluble phenolic metabolites have low thermal stability and contribute to a large proportion of phenolic product yield at low temperature. When comparing the mutant to the wild type, the water treated mutant biomass yield 40% less phenolic products at low temperature than wild type biomass (Figure 4-2A). This indicates soluble phenolic metabolites removed by water did not contribute to the yield reduction in the mutant. After removing both cell wall incorporated HCAs and soluble phenolic metabolites by NaOH, the mutant biomass still yield 30% less phenolic products at low temperature than the wild type biomass (Figure 4-2A). This suggests lignin has a main contribution to the low

temperature yield reduction in the COMT mutant. A comparison between water treated samples and NaOH treated samples indicates HCAs may also contribute to the yield reduction since the reduction was diminished after the removal of HCAs (Figure 4-2A). For high temperature regime and total phenolic product yield, the COMT-3 line shows no significant change comparing to wild type but the COMT-2 line shows a yield reduction (Figure 4-2 B and C). The reason for the yield reduction of COMT2 at both high-temperature and low temperature is unknown. One possibility is some parts of altered lignin become highly thermally unstable and were converted to light gases like $CO_2$ rather than phenolic products even at low temperature.

Overall, these results indicate that switchgrass COMT mutant biomass produce fewer amounts of phenolic products upon low-temperature conversion than their corresponding wild-types. The lignin structure and the amount of cell wall incorporated HCAs may contribute to the level of lignin decomposition at low temperature. Based on a previous NMR study on the lignin of an Arabidopsis COMT mutant, decreased abundance of β–O–4 inter-unit linkages in the mutants is caused by the suppression of ***S*** monomer formation. This structural change can lead to higher bond dissociation enthalpy distributions by reducing percentage of thermally unstable bonds (Table 4-1), and therefore an increase in the overall thermal stability of lignin. Admittedly, there could be other uncharacterized changes in the COMT mutant that also contribute to the low temperature yield reduction. To further confirm S/G ratio is the main contributor,

we will test how the phenolic yield at low temperature influenced by the natural variation of lignin S/G ratio in diverse switchgrass genotypes in the future.

Besides more detailed mechanism study, we also evaluate the fractionated thermal products from COMT mutant can benefit downstream processing. Since COMT-3 biomass have a lower phenolic yield at low temperature and no reduction in total phenolic yield, we would expect better performance of condensation/ketonization catalysts on the low temperature fraction of thermal products and unaltered performance of hydrodeoxygenation on high temperature faction.

## Materials and Methods

### *Biomass Source*

C. N. Stewart and associates generously provided dry switchgrass biomass of two independent Caffeic acid O-methyltransferase knock down lines COMT2 and COMT3, as described and characterized in a previous study(Baxter et al. 2014). The isogenetic wild-types and mutants exhibit significant differences in their S/G ratios, while the total lignin content is only slightly lower in the mutants.

### *Biomass Preparation*

Dry switchgrass biomass was ground with a Thomas Wiley® Mini-Mill with 60-mesh screen. Approximately 100mg each of 5 biological replicates for each genetic line (COMT-2 and COMT-3) were mixed together to create sample pools. From these mixtures, approximately 1mg of biomass was prepared for analysis in a CDS analytical 5250T pyroprobe system, attached to a Shimadzu QP-2010+ GC/MS-FID. The samples were loaded into a fire polished quartz tube with a filler rod below to prevent the samples from falling out of the bottom. The quartz tubes were weighed before and after sample loading, and the difference was taken as the mass of the sample in the tube. The tubes were then transferred into the pyroprobe for rapid screening.

### *Pyrolysis Experiments*

Pyrolysis experiments were done in a ‘double-shot’ manner; the first temperature regime was 290°C with a hold time of 120 seconds, followed by 500°C for 60 seconds. Evolved vapors from the pyroprobe apparatus were transported via transfer lines heated to 300°C to the injection port of a Shimadzu QP-2010+ GC/MS-FID system with a 60m

RTX-1701 column. The resulting ion chromatogram was used to identify significant peaks in the FID chromatogram, which was used for quantification.

Two publications by (Faix et al. 1991) were used as the primary means of compound identification. Each integrated FID peak area was then divided by the total mass of the material in the sample tube, thereby normalizing each experiment to the initial amount of sample fed into the pyroprobe. Identified compounds were assigned into lumps of compounds based on organic functionalities, in a similar manner as described by Dauenhauer et al. (Paulsen et al. 2013). Lumped areas were summed and the mean values of these compound lumps across the technical replicate experiments performed are reported throughout this work as 'yield'. The identified compounds and their lumps as used in this work are listed in table 3.

***NaOH and Water Treatment***

About 60mg of biomass are weighed and extracted with 10ml 2M NaOH and water separately in 15ml tubes. The tubes are rotated on a shaker for 24h at room temperature. NaOH extraction is terminated by adding 2ml concentrated HCl to adjust pH to 2-3. The tubes are centrifuged at maximum speed for 5min to remove supernatant. The pellets are wash twice with deionized water and lyophilized overnight for thermal conversion analysis. Water washed samples go through the same washing steps as a control.

***Analysis of Hydroxycinnamic Acids***

The hydroxycinnamic acids (HCAs) are extracted and analyzed using the previously published method (Bartley et al. 2013). Briefly, biomasses are extracted with 2M NaOH

at room temperature and quantified by a Dionex Ultimate 3000 HPLC system (Thermo Scientific-Dionex) with C18 column and UV detection.

**Figures and Tables**

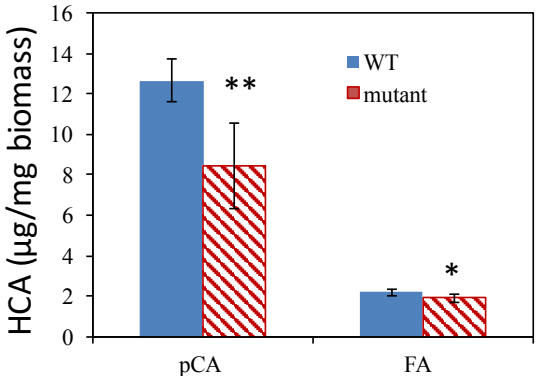



**Figure 4-1**. Hydroxycinnamic acids (HCAs) in switchgrass COMT mutant biomass. Error bars indicate standard deviation of four biological replicates of the COMT2 line. Asterisks indicate the difference is significant at 0.05 level by two tail t-test.

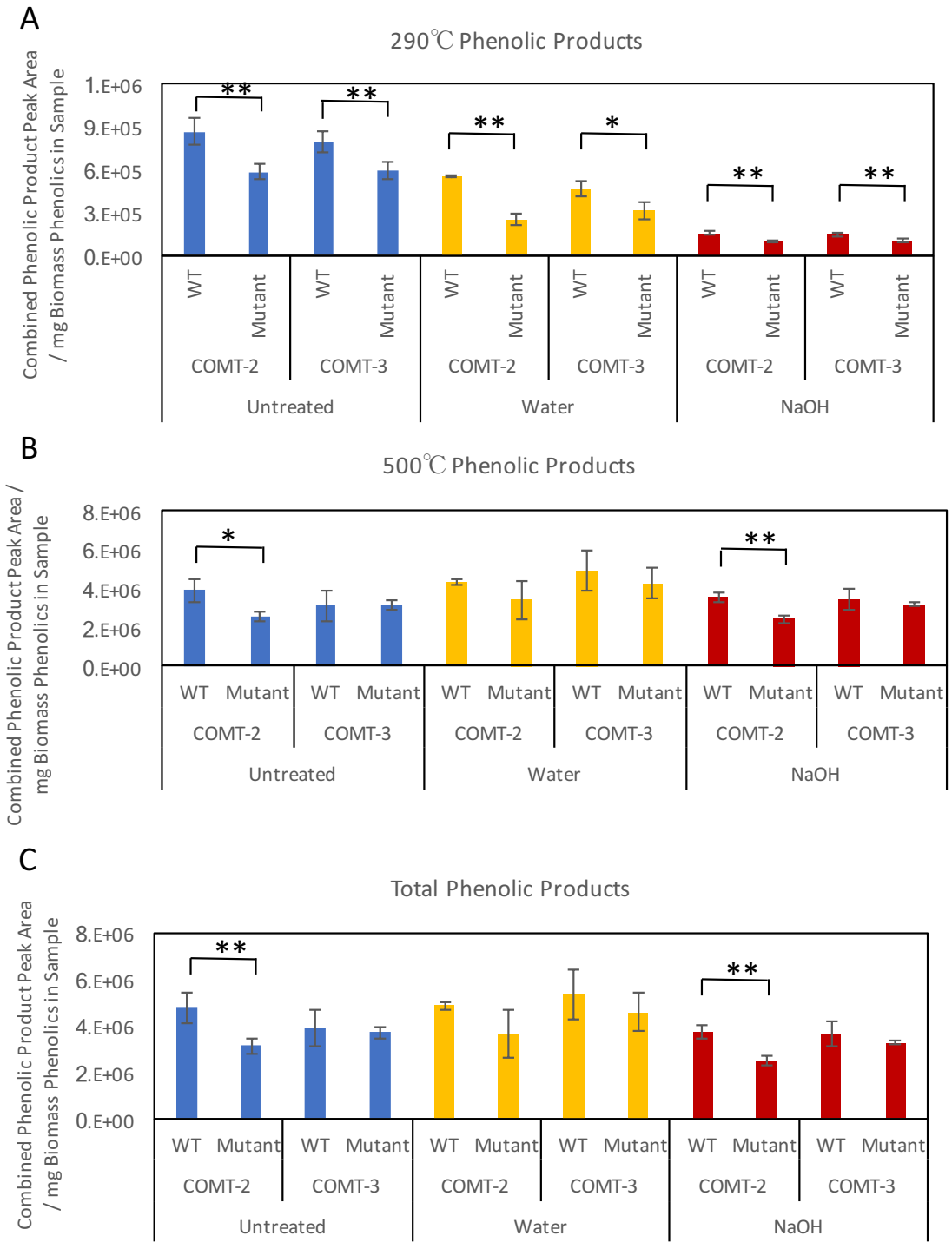

A
290℃ Phenolic Products
Combined Phenolic Product Peak Area / mg Biomass Phenolics in Sample
1.E+06
9.E+05
6.E+05
3.E+05
0.E+00
**
**
**
*
**
**
WT
Mutant
COMT-2
COMT-3
Untreated
Water
NaOH
B
500℃ Phenolic Products
Combined Phenolic Product Peak Area / mg Biomass Phenolics in Sample
8.E+06
6.E+06
4.E+06
2.E+06
0.E+00
*
**
C
Total Phenolic Products
Combined Phenolic Product Peak Area / mg Biomass Phenolics in Sample
8.E+06
6.E+06
4.E+06
2.E+06
0.E+00
**
**

**Figure 4-2** The yields of phenolic products in a two-step thermal conversion at 290°C and 500°C. Yield values are (Summed phenolic product peak areas) / (mg of total biomass phenolics). (A) First step low temperature conversion at 290°C (B) Second step high temperature conversion at 500°C. (C) The summed yield from both steps. The sample for each mutant line is a pool of biomass from 5 plants. Each value reported is the mean of three technical replicates; error bars indicate 95% confidence interval of the mean for these three replicates. Asterisks indicate the difference is significant at 0.05 (*) or 0.01 (**).

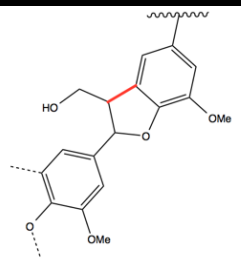
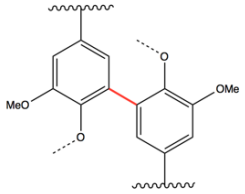
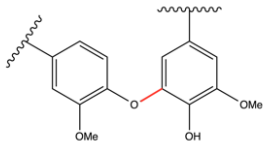
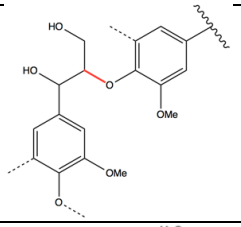
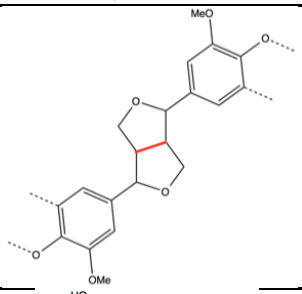
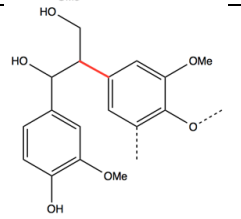
| Bond | Structure | $\Delta H_{diss}$ , kcal/mol, 298K | Reference |
|---|---|---|---|
| **β-5** |  | 101.6 – 107.1 | *(Kim et al., 2011)* |
| **5-5** |  | 111.8 – 118.1 | *(Kim et al., 2011)* |
| **5-O-4** |  | 77.7 – 82.5 | *(Parthasarathi et al., 2011)* |
| **β–O–4** |  | 67.7 – 71.3 | *(Kim et al., 2011)* |
| **β-β** |  | 81.1 – 82.6 | (Elder, 2014) |
| **β-1** |  | 64.7 – 69.1 | *(Parthasarathi et al., 2011)* |

**Table 4-1**. Summarized results from Density Functional Theory (DFT) calculations of the enthalpy of dissociation for six common lignin crosslinks.

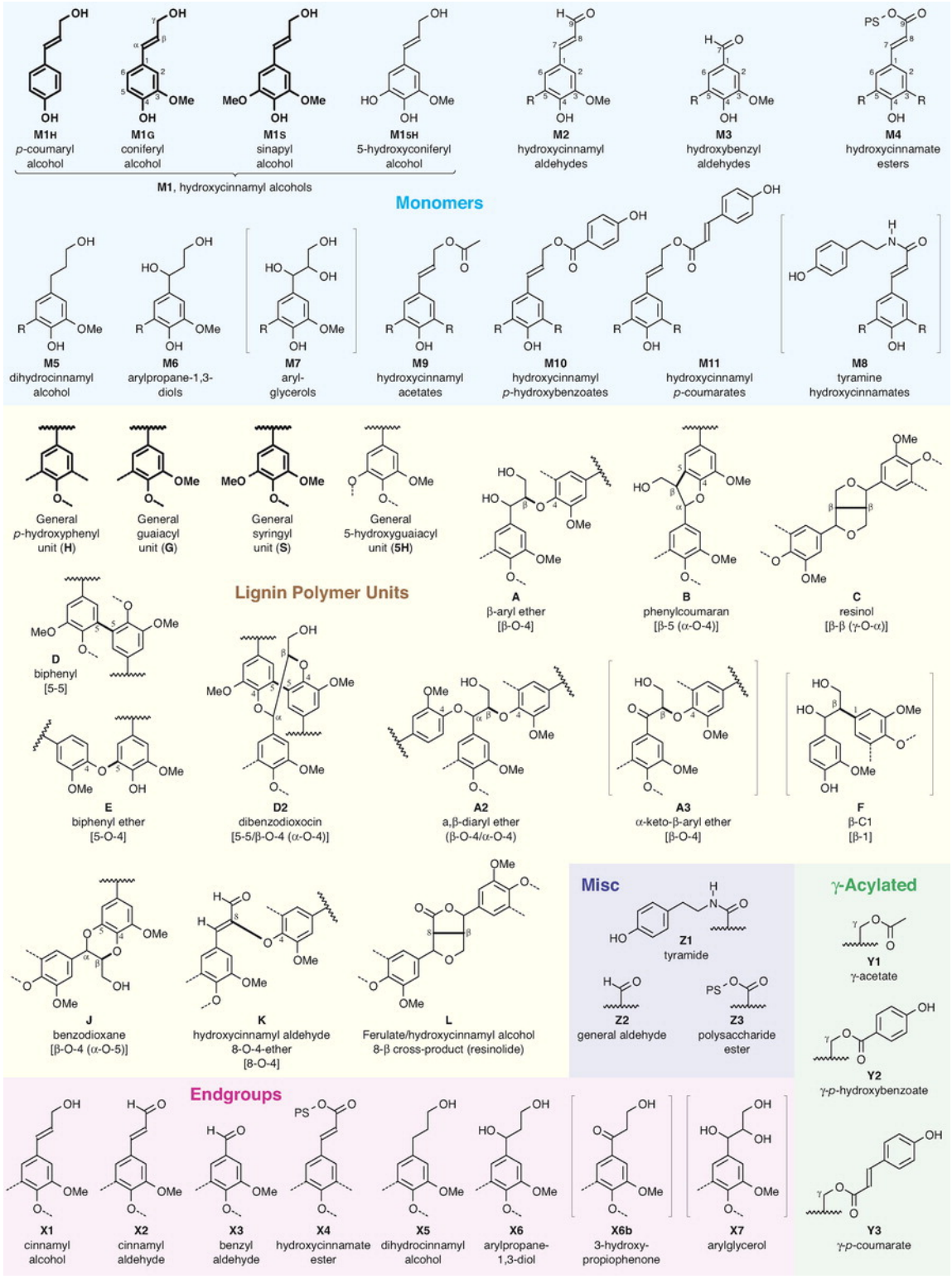



**Figure 4-S1**. Lignin monomers and structures in the polymer. Figure from a review by Boerjan et. al. (Boerjan et. al.2003)

# Chapter 5: Conclusions and Future Directions

With this dissertation, I aimed to expand knowledge of grass cell wall synthesis and provide a proof of concept for how this knowledge could be applied to improve the biomass to biofuel conversion. I achieved three specific objectives as follows: (1) Understanding how cell wall composition changes over development in above-ground organs of rice (2) Identification of candidate cell wall synthesis genes the alteration of which may improve biomass chemical conversion efficiency (3) Demonstration that cell wall engineering can improve thermal fractionation, a means to improve biomass conversion to fuels and other chemical products.

Chapter 2 reported measurements of cell wall composition and transcripts in 30 samples from different rice organs and developmental time points and tested a strategy to identify grass cell wall synthesis genes by correlating cell wall abundance and gene expression. For this strategy, we focused on candidate genes highly expressed in rice instead of looking at the whole transcriptome to reduce the background noise. We compared different correlation methods based on performance of positive and negative controls. We found that the cell wall composition was distinct in young and old tissues, which suggests different synthesis activities in them. Therefore, we subdivided the samples into a young subset and an old subset based on cell wall composition and performed correlation on each subset. With all the efforts to improve signal to noise ratio, we identified nine cell wall candidate genes and corresponding cell wall components. One of the candidate gene, AT9, may incorporate FA into arabinoxylan

and is a promising target to reduce cell wall recalcitrance. However, there is still space to improve since several positive control genes did not correlate with the cell wall components they synthesize. The noise that disrupted the control correlations could be from organ- or tissue-specific expression, post-transcriptional regulations, or substrate avaliablity. Therefore, we decided to focus on a single organ and proteomics study in Chapter 3.

In Chapter 3, we analyzed cell wall composition along eight segments of an elongating stem internode at booting stage and the proteins and metabolites present in the whole stem internode. I choose stems since they represent >50% of the above ground biomass of many species being developed as energy crops (Fu et al. 2011; Olson et al. 2012). Knowledge of cell wall biosynthesis in stem also may be applied to other organs. According to the data from Chapter 2, a majority of cell wall genes expressed in internode samples are also expressed in leaf samples though the magnitude is usually different. Microscopy analysis revealed this elongation internode have uniform tissue structures longitudinally, which we expect to produce less noise caused by the tissue-type changes across different segments. Cell wall analysis reveals active cell wall synthesis, especially in lower segments. We tested several proteomics methods on this elongating internode and end up with 2356 proteins being confidently and repeatedly identified. At this depth, we found most enzymes required for lignin present in this elongating stem. We also identified several GTs and ATs, of which AT9 was the most abundant one by total peptide count. As only 60% of transcripts of the maize-sorghum

syntenic genes are translated to proteins (Walley et al. 2016), the presence of these enzymes suggests they are likely to be functional.

Both Chapter 2 and Chapter 3 were dedicated to developing strategies for studying cell wall biosynthesis with high-throughput technologies (Objective 1). The correlation analysis method developed in Chapter 2 and the proteomics and the cell walls characterization of elongating internodes established in Chapter 3, lays the groundwork for a comprehensive multi-omics experiment to reveal the molecular mechanism of cell wall synthesis and rice stem elongation. More specifically, I propose to measure transcripts, proteins, and phosphopeptides in the eight internode segments corresponding to the eight segments in Chapter 3. By correlating these data with the cell wall measurements in Chapter 3, transcripts, proteins, and protein phosphorylation events that related to cell wall synthesis or remodeling can be identified. Since the internode possesses a unidirectional development gradient, the samples are continuous in development from young to old. This will allow us to interpret the results even if there is a lag between the expression of a cell wall synthesis gene and the accumulation of the cell wall components being synthesized. The other advantage of this experiment is we can interpret the high-throughput measurement together to reveal post-transcriptional regulation and post-translational regulation. Phosphorylation as one important translational-regulation mechanism have been shown to change the activity or the half-life of some cell wall enzymes like CesA (Chen et al. 2010; Taylor 2007), or the activity of cell wall transcription factors such as MYB4 (Morse et al. 2009). A correlation network built on this multi-omics data will not only provide valuable

information about the regulation of cell wall synthesis but also the regulation of other developmental processes like cell division and cell wall elongation that occur in the lower segments (Kende et al. 1998). A recently maize study highlighted that gene regulation networks combining transcriptome, proteome, and phosphoproteome preforms better than the networks with single input (Walley et al. 2016). They evaluated the performance of networks by known target genes of two transcription factors and found the multi-omics network can identify more true positives than single-input networks at the same false positive rate. Therefore, I would expect the multi-omics network have advantages over most current co-expression networks that only built on transcriptomic data (Cao et al. 2012; Mutwil et al. 2011). All these results will guide functional study of cell wall synthesis and regulation and yield genetic alterations that are very likely to improve the efficiency of biofuel production.

The acyltransferase genes and glycosyltransferase genes identified in rice by Project 2 and Project 3 may improve the efficiency of chemical conversion of rice straw, in particular, which is a major agricultural waste globally. The mutants for many of the identified genes such as AT7 and AT9 are available and we are in the process of characterizing them (Jeon et al. 2000). Previous characterization of an AT7 knock-down mutant show a decrease of FA (Bartley et al. 2013), we expect the biomass of this mutant to have elevated sugar yield due to the disruption of FA cross-links. However, the FA phenotype and the potential sugar yield phenotype of this AT7 mutant have to be confirmed in RNAi knockdown lines. We could study the function of AT9 by overexpressing it in Arabidopsis, which only have one AT that was not shown to

influence cell wall associated FA in its mutant. If we are able to confirm the function of AT as an FA transferase in Arabidopsis, we will make knock-down or knock-out rice mutants to see if the sugar yield will be increase or not. The AT genes in switchgrass and other species has been found by a phylogenetic study, which allow the genetic manipulation of them in switchgrass (Karlen et al. 2016).

Chapter 4 demonstrated that COMT mutants with altered lignin structure display better segregation of lignin-derived products from polysaccharide-derived products by yielding less lignin-derived product mass with low temperature thermal treatment than the from the wild type. This indicates the reduction of S subunits in lignin can elevate lignin thermal stability and improve segregation of lignin-derived products from cellulose-derived products. By improving thermal fractionation efficiency and reducing product heterogeneity, downstream processing with catalysts is expected to be more effective. A success on this opens the path to apply other knowledge of cell wall synthesis to improve thermal fractionation.

Besides lignin, some polysaccharides could also be potential targets for improving thermal fractionation. For example, some hemicelluloses, especially those possess complex substituent structures such as xylans, also have relative wide range of decomposition temperature like 250-375°C (Werner et al. 2014). Reducing the higher decomposition temperature of hemicellulose will facilitate better segregation of their products from cellulose-derived products that decompose at 300-350 °C. One thermal conversion experiment on purified xylans indicates unsubstituted xylans have a

decomposition temperature of 250-300°C, which is lower than the decomposition temperature of arabinoxylans at 300-350°C (Werner et al. 2014). The relatively high decomposition temperature of arabinoxylans could due to the FA crosslinks on the arabinose substituent of them. To test this hypothesis, we can reduce FA and arabinose substituents on arabinoxylans by manipulating their sysnthesis genes. Two GT61s involved in the synthesis of the arabinose substituent and the xylose-arabinose substituent have been studied in grasses by genetics (Anders et al. 2012; Chiniquy et al. 2012). An UDP-arabino-mutase gene that synthesizes the precursor of arabinose substituents has been knocked down in switchgrass resulting in a 50% reduction of arabinose in cell walls, which is mostly from arabinoxylans. In Chapter 3, we also identified several putative GT61s that may synthesize xylan substituents and AT9 and AT7, which may incorporate FA into arabinoxylans. An initial thermal fractionation analysis on the rice mutants of these genes will provide a hint. Though this could be tested by chemically modifying the structure of isolated xylans, it is better to retain all other biomass components that closely interact with xylans in real biomass. Due to the complexity of biomass, the interaction among components such covalent cross-links, non-covalent interaction, and the ions that possess catalytic effects also affect thermal conversion (Du et al. 2014; Lin et al. 2015; Zhang et al. 2015). More biomass properties important for thermal fractionation may be identified by measuring biomass components, cell wall features, and thermal fractionation efficiency in natural switchgrass variant and identify associations. More understanding of how thermal fractionation is influenced by biomass properties may further drive basic research on cell wall synthesis or modification.

Since cell walls are involved in biological process such as cell elongation, mechanical support, and pathogen defense, some genetically modified plants suffer biomass yield losses and increased susceptibility to lodging and pathogens (Aohara et al. 2009). This highlights the opportunity and necessity to understand the function of cell walls in plant biology. For example, the the stunted growth of a lignin mutant, *ref8,* is caused by a regulatory pathway activated by lignin deficiency and mediated by the transcriptional activator Mediator complex (Bonawitz et al. 2014). By disrupting the Mediator complex, the lignin mutant shows high sugar yield without developmental defects. Another strategy is to make fine adjustment on the structure of cell wall components to minimize the effect on major biological function. For example, is a OsCESA9 conserved-site mutant have more than 2-fold increase in sugar yield and enhanced lodging resistance by decreasing the degree of polymerization and crystallinity index of cellulose. This work highlights the importance of understanding how to changing the cell wall while retaining or even enhancing cell wall mechanical properties. Another example are the *p*-Coumaroyl-CoA:monolignol transferase mutants. By having more *p*CA esters on lignin, the mutants has greater sugar yields than the wild type without affecting biomass yield (Petrik et al. 2014).

Knowledge of cell wall synthesis obtained in this dissertation will not only advance the efficiency of biomass-biofuel conversion, but also allow us to test the biological functions of cell walls. Some cell wall mutants with altered cell wall composition and

structures influence plant development and pathogen resistance (Bonawitz et al. 2014; Lionetti et al. 2015), while many others have not been examined for these changes in details (Bartley et al. 2013). These mutants can be used to further understand the molecular mechanism, for example, of how plant cells sense deficiencies of cell walls, what proteins are involved in signaling cascades, and how signal pathways cross-talk with other processes such as cell elongation, programed cell death and the synthesis of other metabolites. Better understanding of these effects caused by cell wall changes will allow us to genetically modify grass cell walls without sacrificing important agronomic traits of crops.

## References


Abdi, F., Quinn, J.F., Jankovic, J., McIntosh, M., Leverenz, J.B., Peskind, E., et al. (2006) Detection of biomarkers with a multiplex quantitative proteomic platform in cerebrospinal fluid of patients with neurodegenerative disorders. *J Alzheimers Dis* 9: 293-348.

Abedon, B.G., Hatfield, R.D. and Tracy, W.F. (2006) Cell wall composition in juvenile and adult leaves of maize (Zea mays L.). *J Agr Food Chem* 54: 3896-3900.

Aden A., R.M., Ibsen K. , Jechura J., Neeves K. , Sheehan J. , Wallace B. (2002) Lignocellulosic Biomass to Ethanol Process Design and Economics Utilizing Co-Current Dilute Acid Prehydrolysis and Enzymatic Hydrolysis for Corn Stover. *NREL/TP* 510-32438.

Adjei-Afriyie, F., Kim, C.S., Takemura, M., Ishikawa, M., Tebayashi, S. and Horiike, M. (2000) Probing stimulants from the rice plant towards the smaller brown planthopper, Laodelphax striatellus (FALLEN) (Homoptera: Delphacidae). *Z Naturforsch C* 55: 1038-1043.

Albenne, C., Canut, H. and Jamet, E. (2013) Plant cell wall proteomics: the leadership of Arabidopsis thaliana. *Front Plant Sci* 4: 111.

Alonso-Simon, A., Kristensen, J.B., Obro, J., Felby, C., Willats, W.G.T. and Jorgensen, H. (2010) High-Throughput Microarray Profiling of Cell Wall Polymers During Hydrothermal Pre-Treatment of Wheat Straw. *Biotechnol Bioeng* 105: 509-514.

Anders, N., Wilkinson, M.D., Lovegrove, A., Freeman, J., Tryfona, T., Pellny, T.K., et al. (2012) Glycosyl transferases in family 61 mediate arabinofuranosyl transfer onto xylan in grasses. *Proc. Natl. Acad. Sci. U. S. A.* 109: 989-993.

Antal, M.J., Jr. and Varhegyi, G. (1995) Cellulose Pyrolysis Kinetics: The Current State of Knowledge. *Industrial & Engineering Chemistry Research* 34: 703-717.

Aohara, T., Kotake, T., Kaneko, Y., Takatsuji, H., Tsumuraya, Y. and Kawasaki, S. (2009) Rice BRITTLE CULM 5 (BRITTLE NODE) is Involved in Secondary Cell Wall Formation in the Sclerenchyma Tissue of Nodes. *Plant Cell Physiol* 50: 1886-1897.

Atmodjo, M.A., Hao, Z.Y. and Mohnen, D. (2013) Evolving Views of Pectin Biosynthesis. *Annual Review of Plant Biology, Vol 64* 64: 747-+.

Azuma, T., Sumida, Y., Kaneda, Y., Uchida, N. and Yasuda, T. (1996) Changes in cell wall polysaccharides in the internodes of submerged floating rice. *Plant Growth Regul* 19: 183-187.

Bals, B., Rogers, C., Jin, M., Balan, V. and Dale, B. (2010) Evaluation of ammonia fibre expansion (AFEX) pretreatment for enzymatic hydrolysis of switchgrass harvested in different seasons and locations. *Biotechnol Biofuels* 3: 1.

Bandong, J.P. and Litsinger, J.A. (2005) Rice crop stage susceptibility to the rice yellow stemborer Scirpophaga incertulas (Walker) (Lepidoptera : Pyralidae). *Int. J. Pest Manag.* 51: 37-43.

Barnett, J.R. and Bonham, V.A. (2004) Cellulose microfibril angle in the cell wall of wood fibres. *Biol Rev* 79: 461-472.

Bartley, L.E., Peck, M.L., Kim, S.R., Ebert, B., Maniseri, C., Chiniquy, D., et al. (2013) Overexpression of a BAHD Acyltransferase, OsAt10, alters rice cell wall hydroxycinnamic acid content and saccharification. *Plant Physiol*.

Bashline, L., Lei, L., Li, S.D. and Gu, Y. (2014a) Cell Wall, Cytoskeleton, and Cell Expansion in Higher Plants. *Mol Plant* 7: 586-600.

Bashline, L., Li, S. and Gu, Y. (2014b) The trafficking of the cellulose synthase complex in higher plants. *Ann Bot* 114: 1059-1067.

Baskin, T.I. (2005) Anisotropic expansion of the plant cell wall. *Annu Rev Cell Dev Bi* 21: 203-222.

Baxter, H.L., Mazarei, M., Labbe, N., Kline, L.M., Cheng, Q.K., Windham, M.T., et al. (2014) Two-year field analysis of reduced recalcitrance transgenic switchgrass. *Plant Biotechnol J* 12: 914-924.

Beauzamy, L., Louveaux, M., Hamant, O. and Boudaoud, A. (2015) Mechanically, the Shoot Apical Meristem of Arabidopsis Behaves like a Shell Inflated by a Pressure of About 1 MPa. *Front Plant Sci* 6.

Boerjan, W., Ralph, J. and Baucher, M. (2003) Lignin biosynthesis. *Annu. Rev. Plant Biol.* 54: 519-546.

Boller, T. and Felix, G. (2009) A Renaissance of Elicitors: Perception of Microbe-Associated Molecular Patterns and Danger Signals by Pattern-Recognition Receptors. *Annu. Rev. Plant Biol.* 60: 379-406.

Bonawitz, N.D., Kim, J.I., Tobimatsu, Y., Ciesielski, P.N., Anderson, N.A., Ximenes, E., et al. (2014) Disruption of Mediator rescues the stunted growth of a lignin-deficient Arabidopsis mutant. *Nature* 509: 376-380.

Bosch, M., Mayer, C.D., Cookson, A. and Donnison, I.S. (2011) Identification of genes involved in cell wall biogenesis in grasses by differential gene expression profiling of elongating and non-elongating maize internodes. *J Exp Bot* 62: 3545-3561.

Bouaziz, M., Simmonds, M.S., Grayer, R.J., Kite, G.C. and Damak, M. (2001) Flavonoids from Hyparrhenia hirta Stapf (Poaceae) growing in Tunisia. *Biochem. Syst. Ecol.* 29: 849-851.

Bowles, D., Isayenkova, J., Lim, E.K. and Poppenberger, B. (2005) Glycosyltransferases: managers of small molecules. *Curr Opin Plant Biol* 8: 254-263.

Bridgwater, A.V. (1994) Catalysis in Thermal Biomass Conversion. *Appl Catal a-Gen* 116: 5-47.

Brutus, A., Sicilia, F., Macone, A., Cervone, F. and De Lorenzo, G. (2010) A domain swap approach reveals a role of the plant wall-associated kinase 1 (WAK1) as a receptor of oligogalacturonides. *P Natl Acad Sci USA* 107: 9452-9457.

Buanafina, M.M.D. (2009) Feruloylation in grasses: current and future perspectives. *Mol Plant* 2: 861-872.

Buanafina, M.M.D., Fescemyer, H.W., Sharma, M. and Shearer, E.A. (2016) Functional testing of a PF02458 homologue of putative rice arabinoxylan feruloyl transferase genes in Brachypodium distachyon. *Planta* 243: 659-674.

Buckeridge, M.S., Rayon, C., Urbanowicz, B., Tine, M.A.S. and Carpita, N.C. (2004) Mixed linkage (1 -> 3),(1 -> 4)-beta-D-glucans of grasses. *Cereal Chem* 81: 115-127.

Bunzel, M., Ralph, J., Lu, F., Hatfield, R.D. and Steinhart, H. (2004) Lignins and ferulate-coniferyl alcohol cross-coupling products in cereal grains. *J Agr Food Chem* 52: 6496-6502.

Burton, R.A. and Fincher, G.B. (2012) Current challenges in cell wall biology in the cereals and grasses. *Front Plant Sci* 3.

Burton, R.A., Jobling, S.A., Harvey, A.J., Shirley, N.J., Mather, D.E., Bacic, A., et al. (2008) The genetics and transcriptional profiles of the cellulose synthase-like HvCslF gene family in barley. *Plant Physiol* 146: 1821-1833.

Burton, R.A., Wilson, S.M., Hrmova, M., Harvey, A.J., Shirley, N.J., Medhurst, A., et al. (2006) Cellulose synthase-like CslF genes mediate the synthesis of cell wall (1,3;1,4)-beta-D-glucans. *Science* 311: 1940-1942.

Busse-Wicher, M., Grantham, N.J., Lyczakowski, J.J., Nikolovski, N. and Dupree, P. (2016) Xylan decoration patterns and the plant secondary cell wall molecular architecture. *Biochem. Soc. Trans.* 44: 74-78.

Caffall, K.H., Pattathil, S., Phillips, S.E., Hahn, M.G. and Mohnen, D. (2009) Arabidopsis thaliana T-DNA Mutants Implicate GAUT Genes in the Biosynthesis of Pectin and Xylan in Cell Walls and Seed Testa. *Mol Plant* 2: 1000-1014.

Calderan-Rodrigues, M.J., Jamet, E., Douche, T., Bonassi, M.B.R., Cataldi, T.R., Fonseca, J.G., et al. (2016) Cell wall proteome of sugarcane stems: comparison of a destructive and a non-destructive extraction method showed differences in glycoside hydrolases and peroxidases. *BMC Plant Biol.* 16.

Cao, P.J., Bartley, L.E., Jung, K.H. and Ronald, P.C. (2008) Construction of a rice glycosyltransferase phylogenomic database and identification of rice-diverged glycosyltransferases. *Mol Plant* 1: 858-877.

Cao, P.J., Jung, K.H., Choi, D., Hwang, D., Zhu, J. and Ronald, P.C. (2012) The Rice Oligonucleotide Array Database: an atlas of rice gene expression. *Rice* 5.

Carlson, T.R., Tompsett, G.A., Conner, W.C. and Huber, G.W. (2009) Aromatic Production from Catalytic Fast Pyrolysis of Biomass-Derived Feedstocks. *Top Catal* 52: 241-252.

Carlson, T.R., Vispute, T.R. and Huber, G.W. (2008) Green gasoline by catalytic fast pyrolysis of solid biomass derived compounds. *Chemsuschem* 1: 397-400.

Carnelli, A.L., Madella, M. and Theurillat, J.P. (2001) Biogenic silica production in selected alpine plant species and plant communities. *Annals of Botany* 87: 425-434.

Carpita, N.C. (1989) Pectic Polysaccharides of Maize Coleoptiles and Proso Millet Cells in Liquid Culture. *Phytochemistry* 28: 121-125.

Carpita, N.C. (1996) Structure and biogenesis of the cell walls of grasses. *Annu Rev Plant Phys* 47: 445-476.

Chateigner-Boutin, A.L., Ordaz-Ortiz, J.J., Alvarado, C., Bouchet, B., Durand, S., Verhertbruggen, Y., et al. (2016) Developing pericarp of maize: a model to study arabinoxylan synthesis and feruloylation. *Front Plant Sci* 7: 1476.

Chen, S., Ehrhardt, D.W. and Somerville, C.R. (2010) Mutations of cellulose synthase (CESA1) phosphorylation sites modulate anisotropic cell expansion and bidirectional mobility of cellulose synthase. *Proc. Natl. Acad. Sci. U. S. A.* 107: 17188-17193.

Chen, S.L., Jia, H.L., Zhao, H.Y., Liu, D., Liu, Y.M., Liu, B.Y., et al. (2016) Anisotropic cell expansion is affected through the bidirectional mobility of cellulose synthase complexes and phosphorylation at two critical residues on CESA3. *Plant Physiol* 171: 242-250.

Chen, X.W., Vega-Sanchez, M.E., Verhertbruggen, Y., Chiniquy, D., Canlas, P.E., Fagerstrom, A., et al. (2013) Inactivation of OsIRX10 Leads to Decreased Xylan Content in Rice Culm Cell Walls and Improved Biomass Saccharification. *Mol Plant* 6: 570-573.

Chen, X.Y., Kim, S.T., Cho, W.K., Rim, Y., Kim, S., Kim, S.W., et al. (2009) Proteomics of weakly bound cell wall proteins in rice calli. *J. Plant Physiol.* 166: 675-685.

Cheng, S.Y., Wei, L., Zhao, X.H. and Julson, J. (2016) Application, Deactivation, and Regeneration of Heterogeneous Catalysts in Bio-Oil Upgrading. *Catalysts* 6.

Chiniquy, D., Sharma, V., Schultink, A., Baidoo, E.E., Rautengarten, C., Cheng, K., et al. (2012) XAX1 from glycosyltransferase family 61 mediates xylosyltransfer to rice xylan. *P Natl Acad Sci USA* 109: 17117-17122.

Chiniquy, D., Varanasi, P., Oh, T., Harholt, J., Katnelson, J., Singh, S., et al. (2013) Three novel rice genes closely related to the Arabidopsis IRX9, IRX9L, and IRX14 genes and their roles in xylan biosynthesis. *Front Plant Sci* 4.

Cho, J.Y., Moon, J.H., Seong, K.Y. and Park, K.H. (1998) Antimicrobial activity of 4-hydroxybenzoic acid and trans 4-hydroxycinnamic acid isolated and identified from rice hull. *Biosci Biotech Bioch* 62: 2273-2276.

Cho, W.K., Chen, X.Y., Chu, H., Rim, Y., Kim, S., Kim, S.T., et al. (2009) Proteomic analysis of the secretome of rice calli. *Physiol. Plant* 135: 331-341.

Christensen, U., Alonso-Simon, A., Scheller, H.V., Willats, W.G. and Harholt, J. (2010) Characterization of the primary cell walls of seedlings of Brachypodium distachyon--a potential model plant for temperate grasses. *Phytochemistry* 71: 62-69.

Chuck, G.S., Tobias, C., Sun, L., Kraemer, F., Li, C.L., Dibble, D., et al. (2011) Overexpression of the maize Corngrass1 microRNA prevents flowering, improves digestibility, and increases starch content of switchgrass. *P Natl Acad Sci USA* 108: 17550-17555.

Chung, D., Pattathil, S., Biswal, A.K., Hahn, M.G., Mohnen, D. and Westpheling, J. (2014) Deletion of a gene cluster encoding pectin degrading enzymes in Caldicellulosiruptor bescii reveals an important role for pectin in plant biomass recalcitrance. *Biotechnology for biofuels* 7.

Cosgrove, D.J. (2000) Loosening of plant cell walls by expansins. *Nature* 407: 321-326.

Cosgrove, D.J. (2005) Growth of the plant cell wall. *Nat Rev Mol Cell Bio* 6: 850-861.

Cosgrove, D.J. (2016a) Catalysts of plant cell wall loosening. *F1000Res* 5.

Cosgrove, D.J. (2016b) Plant cell wall extensibility: connecting plant cell growth with cell wall structure, mechanics, and the action of wall-modifying enzymes. *J Exp Bot* 67: 463-476.

Costa, R.M.F., Lee, S.J., Allison, G.G., Hazen, S.P., Winters, A. and Bosch, M. (2014) Genotype, development and tissue-derived variation of cell-wall properties in the lignocellulosic energy crop Miscanthus. *Annals of Botany* 114: 1265-1277.

Cui, K., He, C.Y., Zhang, J.G., Duan, A.G. and Zeng, Y.F. (2012) Temporal and spatial profiling of internode elongation-associated protein expression in rapidly growing culms of bamboo. *J. Proteome Res.* 11: 2492-2507.

Dardick, C., Chen, J., Richter, T., Ouyang, S. and Ronald, P. (2007) The rice kinase database. A phylogenomic database for the rice kinome. *Plant Physiol* 143: 579-586.

Dardick, C. and Ronald, P. (2006) Plant and animal pathogen recognition receptors signal through non-RD kinases. *Plos Pathog* 2: e2.

de Melo, S.P., Monteiro, F.A. and De Bona, F.D. (2010) Silicon distribution and accumulation in shoot tissue of the tropical forage grass Brachiaria brizantha. *Plant Soil* 336: 241-249.

de Oliveira, D.M., Finger-Teixeira, A., Mota, T.R., Salvador, V.H., Moreira-Vilar, F.C., Molinari, H.B.C., et al. (2015) Ferulic acid: a key component in grass lignocellulose recalcitrance to hydrolysis. *Plant Biotechnol J* 13: 1224-1232.

De Souza, A.P., Kamei, C.L.A., Torres, A.F., Pattathil, S., Hahn, M.G., Trindade, L.M., et al. (2015) How cell wall complexity influences saccharification efficiency in Miscanthus sinensis. *J Exp Bot* 66: 4351-4365.

del Rio, J.C., Rencoret, J., Prinsen, P., Martinez, A.T., Ralph, J. and Gutierrez, A. (2012) Structural Characterization of Wheat Straw Lignin as Revealed by Analytical Pyrolysis, 2D-NMR, and Reductive Cleavage Methods. *J Agr Food Chem* 60: 5922-5935.

DeMartini, J.D., Pattathil, S., Miller, J.S., Li, H.J., Hahn, M.G. and Wyman, C.E. (2013) Investigating plant cell wall components that affect biomass recalcitrance in poplar and switchgrass. *Energ Environ Sci* 6: 898-909.

Distler, U., Kuharev, J., Navarro, P., Levin, Y., Schild, H. and Tenzer, S. (2014) Drift time-specific collision energies enable deep-coverage data-independent acquisition proteomics. *Nat Methods* 11: 167-170.

Dixon, R.A. (2001) Natural products and plant disease resistance. *Nature* 411: 843-847.

Djakovic-Petrovic, T., de Wit, M., Voesenek, L.A. and Pierik, R. (2007) DELLA protein function in growth responses to canopy signals. *Plant J* 51: 117-126.

Doblin, M.S., Pettolino, F.A., Wilson, S.M., Campbell, R., Burton, R.A., Fincher, G.B., et al. (2009) A barley cellulose synthase-like CSLH gene mediates (1,3;1,4)-beta-D-glucan synthesis in transgenic Arabidopsis. *Proc. Natl. Acad. Sci. U. S. A.* 106: 5996-6001.

Doering, A., Lathe, R. and Persson, S. (2012) An Update on Xylan Synthesis. *Mol Plant*.

Dong, X.Y., Hong, Z.L., Chatterjee, J., Kim, S.H. and Verma, D.P.S. (2008) Expression of callose synthase genes and its connection with Npr1 signaling pathway during pathogen infection. *Planta* 229: 87-98.

Douche, T., Clemente, H.S., Burlat, V., Roujol, D., Valot, B., Zivy, M., et al. (2013) Brachypodium distachyon as a model plant toward improved biofuel crops: Search for secreted proteins involved in biogenesis and disassembly of cell wall polymers. *Proteomics* 13: 2438-2454.

Du, X.Y., Perez-Boada, M., Fernandez, C., Rencoret, J., del Rio, J.C., Jimenez-Barbero, J., et al. (2014) Analysis of lignin-carbohydrate and lignin-lignin linkages after hydrolase treatment of xylan-lignin, glucomannan-lignin and glucan-lignin complexes from spruce wood. *Planta* 239: 1079-1090.

Eddy, R., Acosta, K., Liu, Y. and Russell, M. (2016) Optimizing greenhouse rice production: what is the best pot size? *Purdue Methods for Rice Growth*: 9.

Egelund, J., Petersen, B.L., Motawia, M.S., Damager, I., Faik, A., Olsen, C.E., et al. (2006) Arabidopsis thaliana RGXT1 and RGXT2 encode Golgi-localized (1,3)-alpha-D-xylosyltransferases involved in the synthesis of pectic rhamnogalacturonan-II. *Plant Cell* 18: 2593-2607.

Faik, A. (2010) Xylan Biosynthesis: News from the Grass. *Plant Physiol* 153: 396-402.

Faix, O., Fortmann, I., Bremer, J. and Meier, D. (1991) Thermal-Degradation Products of Wood - a Collection of Electron-Impact (Ei) Mass-Spectra of Polysaccharide Derived Products. *Holz Roh Werkst* 49: 299-304.

Feller, A., Machemer, K., Braun, E.L. and Grotewold, E. (2011) Evolutionary and comparative analysis of MYB and bHLH plant transcription factors. *Plant J* 66: 94-116.

Ferre, H., Broberg, A., Duus, J.O. and Thomsen, K.K. (2000) A novel type of arabinoxylan arabinofuranohydrolase isolated from germinated barley analysis of substrate preference and specificity by nano-probe NMR. *Eur. J. Biochem.* 267: 6633-6641.

Fincher, G.B. (2009) Revolutionary times in our understanding of cell wall biosynthesis and remodeling in the grasses. *Plant Physiol* 149: 27-37.

Fornale, S., Rencoret, J., Garcia-Calvo, L., Encina, A., Rigau, J., Gutierrez, A., et al. (2017) Changes in Cell Wall Polymers and Degradability in Maize Mutants Lacking 3 ' -and 5 '-O-Methyltransferases Involved in Lignin Biosynthesis. *Plant Cell Physiol* 58: 240-255.

Frankova, L. and Fry, S.C. (2013) Biochemistry and physiological roles of enzymes that cut and paste plant cell-wall polysaccharides. *J Exp Bot* 64: 3519-3550.

Fraser, C.M. and Chapple, C. (2011) The phenylpropanoid pathway in Arabidopsis. *Arabidopsis Book* 9: e0152.

Freshour, G., Clay, R.P., Fuller, M.S., Albersheim, P., Darvill, A.G. and Hahn, M.G. (1996) Developmental and tissue-specific structural alterations of the cell-wall polysaccharides of Arabidopsis thaliana roots. *Plant Physiol* 110: 1413-1429.

Fu, C., Mielenz, J.R., Xiao, X., Ge, Y., Hamilton, C.Y., Rodriguez, M., et al. (2011) Genetic manipulation of lignin reduces recalcitrance and improves ethanol production from switchgrass. *Proceedings of the National Academy of Sciences* 108: 3803-3808.

Fu, C.X., Sunkar, R., Zhou, C.E., Shen, H., Zhang, J.Y., Matts, J., et al. (2012) Overexpression of miR156 in switchgrass (Panicum virgatum L.) results in various morphological alterations and leads to improved biomass production. *Plant Biotechnol J* 10: 443-452.

Fukushima, R.S. and Hatfield, R.D. (2004) Comparison of the acetyl bromide spectrophotometric method with other analytical lignin methods for determining lignin concentration in forage samples. *J Agr Food Chem* 52: 3713-3720.

Gardner, R.O. (1977) Systematic distribution and ecological function of secondary metabolites of Rosidae-Asteridae. *Biochem. Syst. Ecol.* 5: 29-35.

Gershenzon, J. and Dudareva, N. (2007) The function of terpene natural products in the natural world. *Nat. Chem. Biol.* 3: 408-414.

Gibeaut, D.M., Pauly, M., Bacic, A. and Fincher, G.B. (2005) Changes in cell wall polysaccharides in developing barley (Hordeum vulgare) coleoptiles. *Planta* 221: 729-738.

Gilbert, H.J. (2010) The Biochemistry and Structural Biology of Plant Cell Wall Deconstruction. *Plant Physiol* 153: 444-455.

Gille, S., Hansel, U., Ziemann, M. and Pauly, M. (2009) Identification of plant cell wall mutants by means of a forward chemical genetic approach using hydrolases. *Proc. Natl. Acad. Sci. U. S. A.* 106: 14699-14704.

Glassop, D., Roessner, U., Bacic, A. and Bonnett, G.D. (2007) Changes in the sugarcane metabolome with stem development. Are they related to sucrose accumulation? *Plant Cell Physiol* 48: 573-584.

Goldberg, R., Morvan, C. and Roland, J.C. (1986) Composition, Properties and Localization of Pectins in Young and Mature Cells of the Mung Bean Hypocotyl. *Plant Cell Physiol* 27: 417-429.

Gomez-Vasquez, R., Day, R., Buschmann, H., Randles, S., Beeching, J.R. and Cooper, R.M. (2004) Phenylpropanoids, phenylalanine ammonia lyase and peroxidases in elicitor-challenged cassava (Manihot esculenta) suspension cells and leaves. *Annals of Botany* 94: 87-97.

Goujon, T., Sibout, R., Pollet, B., Maba, B., Nussaume, L., Bechtold, N., et al. (2003) A new Arabidopsis thaliana mutant deficient in the expression of O-methyltransferase impacts lignins and sinapoyl esters. *Plant Mol Biol* 51: 973-989.

Grabber, J.H., Ralph, J., Lapierre, C. and Barriere, Y. (2004) Genetic and molecular basis of grass cell-wall degradability. I. Lignin-cell wall matrix interactions. *Comptes rendus biologies* 327: 455-465.

Gray, J., Caparros-Ruiz, D. and Grotewold, E. (2012) Grass phenylpropanoids: Regulate before using! *Plant Sci* 184: 112-120.

Gritsch, C.S. and Murphy, R.J. (2005) Ultrastructure of fibre and parenchyma cell walls during early stages of culm development in Dendrocalamus asper. *Annals of Botany* 95: 619-629.

Gui, J., Shen, J. and Li, L. (2011) Functional characterization of evolutionarily divergent 4-coumarate:coenzyme a ligases in rice. *Plant Physiol* 157: 574-586.

Guntzer, F., Keller, C. and Meunier, J.D. (2010) Determination of the silicon concentration in plant material using Tiron extraction. *New Phytologist* 188: 902-906.

Guo, D., Chen, F., Inoue, K., Blount, J.W. and Dixon, R.A. (2001) Downregulation of caffeic acid 3-O-methyltransferase and caffeoyl CoA 3-O-methyltransferase in transgenic alfalfa. impacts on lignin structure and implications for the biosynthesis of G and S lignin. *Plant Cell* 13: 73-88.

Guo, K., Zou, W., Feng, Y., Zhang, M., Zhang, J., Tu, F., et al. (2014) An integrated genomic and metabolomic framework for cell wall biology in rice. *BMC genomics* 15: 596.

Haider, S. and Pal, R. (2013) Integrated analysis of transcriptomic and proteomic data. *Curr. Genomics* 14: 91-110.

Haigler, C.H., Betancur, L., Stiff, M.R. and Tuttle, J.R. (2012) Cotton fiber: a powerful single-cell model for cell wall and cellulose research. *Front Plant Sci* 3.

Handakumbura, P.P. and Hazen, S.P. (2012) Transcriptional regulation of grass secondary cell wall biosynthesis: playing catch-up with Arabidopsis thaliana. *Front Plant Sci* 3: 74.

Hara, Y., Yokoyama, R., Osakabe, K., Toki, S. and Nishitani, K. (2014) Function of xyloglucan endotransglucosylase/hydrolases in rice. *Annals of Botany* 114: 1309-1318.

Harholt, J., Suttangkakul, A. and Scheller, H.V. (2010) Biosynthesis of Pectin. *Plant Physiol* 153: 384-395.

Hatfield, R.D., Marita, J.M. and Frost, K. (2008) Characterization of p-coumarate accumulation, p-coumaroyl transferase, and cell wall changes during the development of corn stems. *J Sci Food Agr* 88: 2529-2537.

Helmy, M., Sugiyama, N., Tomita, M. and Ishihama, Y. (2012) The rice proteogenomics database OryzaPG-DB: development, expansion, and new features. *Front Plant Sci* 3: 65.

Hematy, K., Sado, P.E., Van Tuinen, A., Rochange, S., Desnos, T., Balzergue, S., et al. (2007) A receptor-like kinase mediates the response of Arabidopsis cells to the inhibition of cellulose synthesis. *Curr Biol* 17: 922-931.

Herron, J.A., Vann, T., Duong, N., Resasco, D.E., Crossley, S., Lobban, L.L., et al. (2017) A Systems-Level Roadmap for Biomass Thermal Fractionation and Catalytic Upgrading Strategies. *Energy Technol-Ger* 5: 130-150.

Himmel, M.E. and Bayer, E.A. (2009) Lignocellulose conversion to biofuels: current challenges, global perspectives. *Curr Opin Biotech* 20: 316-317.

Hirano, K., Aya, K., Morinaka, Y., Nagamatsu, S., Sato, Y., Antonio, B.A., et al. (2013) Survey of genes involved in rice secondary cell wall formation through a co-expression network. *Plant Cell Physiol* 54: 1803-1821.

Hu, G., Burton, C., Hong, Z. and Jackson, E. (2014) A mutation of the cellulose-synthase-like (CslF6) gene in barley (Hordeum vulgare L.) partially affects the β-glucan content in grains. *J Cereal Sci* 59: 189-195.

Humphreys, J.M. and Chapple, C. (2002) Rewriting the lignin roadmap. *Curr Opin Plant Biol* 5: 224-229.

Ishii, T. (1997) Structure and functions of feruloylated polysaccharides. *Plant Sci* 127: 111-127.

Ishii, T. and Hiroi, T. (1990) Linkage of Phenolic-Acids to Cell-Wall Polysaccharides of Bamboo Shoot. *Carbohyd Res* 206: 297-310.

Itoh, J.-I., Nonomura, K.-I., Ikeda, K., Yamaki, S., Inukai, Y., Yamagishi, H., et al. (2005) Rice Plant Development: from Zygote to Spikelet. *Plant Cell Physiol* 46: 23-47.

Jain, M., Nijhawan, A., Tyagi, A.K. and Khurana, J.P. (2006) Validation of housekeeping genes as internal control for studying gene expression in rice by quantitative real-time PCR. *Biochem. Biophys. Res. Commun.* 345: 646-651.

Jamet, E., Canut, H., Boudart, G. and Pont-Lezica, R.F. (2006) Cell wall proteins: a new insight through proteomics. *Trends Plant Sci* 11: 33-39.

Jensen, J.K., Johnson, N.R. and Wilkerson, C.G. (2014) Arabidopsis thaliana IRX10 and two related proteins from psyllium and Physcomitrella patens are xylan xylosyltransferases. *Plant J* 80: 207-215.

Jeon, J.S., Lee, S., Jung, K.H., Jun, S.H., Jeong, D.H., Lee, J., et al. (2000) T-DNA insertional mutagenesis for functional genomics in rice. *Plant J* 22: 561-570.

Johnson, K.L., Jones, B.J., Bacic, A. and Schultz, C.J. (2003) The fasciclin-like arabinogalactan proteins of arabidopsis. A multigene family of putative cell adhesion molecules. *Plant Physiol* 133: 1911-1925.

Jung, H.G. and Allen, M.S. (1995) Characteristics of Plant-Cell Walls Affecting Intake and Digestibility of Forages by Ruminants. *J Anim Sci* 73: 2774-2790.

Jung, Y.H., Jeong, S.H., Kim, S.H., Singh, R., Lee, J.E., Cho, Y.S., et al. (2008) Systematic Secretome Analyses of Rice Leaf and Seed Callus Suspension-Cultured Cells: Workflow Development and Establishment of High-Density Two-Dimensional Gel Reference Maps. *J. Proteome Res.* 7: 5187-5210.

Karlen, S.D., Zhang, C.C., Peck, M.L., Smith, R.A., Padmakshan, D., Helmich, K.E., et al. (2016) Monolignol ferulate conjugates are naturally incorporated into plant lignins. *Sci Adv* 2.

Kato, Y. and Nevins, D.J. (1992) Structural characterization of an arabinoxylan-rhamnogalacturonan complex from cell walls of Zea shoots. *Carbohydrate research* 227: 315-329.

Kawasaki, T., Koita, H., Nakatsubo, T., Hasegawa, K., Wakabayashi, K., Takahashi, H., et al. (2006) Cinnamoyl-CoA reductase, a key enzyme in lignin biosynthesis, is an effector of small GTPase Rac in defense signaling in rice. *P Natl Acad Sci USA* 103: 230-235.

Keegstra, K. (2010) Plant Cell Walls. *Plant Physiol* 154: 483-486.

Kellogg, E.A. (2001) Evolutionary history of the grasses. *Plant Physiol* 125: 1198-1205.

Kelly, R.T., Page, J.S., Luo, Q.Z., Moore, R.J., Orton, D.J., Tang, K.Q., et al. (2006) Chemically etched open tubular and monolithic emitters for nanoelectrospray ionization mass spectrometry. *Anal. Chem.* 78: 7796-7801.

Kende, H., van der Knaap, E. and Cho, H.T. (1998) Deepwater rice: A model plant to study stem elongation. *Plant Physiol* 118: 1105-1110.

Kim, I.A., Kim, B.G., Kim, M. and Ahn, J.H. (2012) Characterization of hydroxycinnamoyltransferase from rice and its application for biological synthesis of hydroxycinnamoyl glycerols. *Phytochemistry* 76: 25-31.

Kim, S. and Pevzner, P.A. (2014) MS-GF plus makes progress towards a universal database search tool for proteomics. *Nat Commun* 5.

Kim, W.D., Kobayashi, O., Kaneko, S., Sakakibara, Y., Park, G.G., Kusakabe, I., et al. (2002) alpha-Galactosidase from cultured rice (Oryza sativa L. var. Nipponbare) cells. *Phytochemistry* 61: 621-630.

Kimball Christensen, A.S. (2008) Does the use of Cannabis species for the production of biodiesel and ethanol, result in higher yields of ethanol than competing cellulotic crops, including Zea mays.

Kodama, O., Suzuki, T., Miyakawa, J. and Akatsuka, T. (1988) Ultraviolet-induced accumulation of phytoalexins in rice leaves. *Agric. Biol. Chem.* 52: 2469-2473.

Koller, A., Washburn, M.P., Lange, B.M., Andon, N.L., Deciu, C., Haynes, P.A., et al. (2002) Proteomic survey of metabolic pathways in rice. *P Natl Acad Sci USA* 99: 11969-11974.

Konishi, T., Aohara, T., Igasaki, T., Hayashi, N., Miyazaki, Y., Takahashi, A., et al. (2011) Down-regulation of UDP-arabinopyranose mutase reduces the proportion of arabinofuranose present in rice cell walls. *Phytochemistry* 72: 1962-1968.

Konishi, T., Miyazaki, Y., Yamakawa, S., Iwai, H., Satoh, S. and Ishii, T. (2010) Purification and Biochemical Characterization of Recombinant Rice UDP-Arabinopyranose Mutase Generated in Insect Cells. *Biosci Biotech Bioch* 74: 191-194.

Konishi, T., Takeda, T., Miyazaki, Y., Ohnishi-Kameyama, M., Hayashi, T., O'Neill, M.A., et al. (2007) A plant mutase that interconverts UDP-arabinofuranose and UDP-arabinopyranose. *Glycobiology* 17: 345-354.

Kozlova, L.V., Ageeva, M.V., Ibragimova, N.N. and Gorshkova, T.A. (2014) Arrangement of mixed-linkage glucan and glucuronoarabinoxylan in the cell walls of growing maize roots. *Annals of Botany* 114: 1135-1145.

Krishnan, A., Guiderdoni, E., An, G., Hsing, Y.I.C., Han, C.D., Lee, M.C., et al. (2009) Mutant Resources in Rice for Functional Genomics of the Grasses. *Plant Physiol* 149: 165-170.

Lairson, L.L., Henrissat, B., Davies, G.J. and Withers, S.G. (2008) Glycosyltransferases: Structures, functions, and mechanisms. *Annu. Rev. Biochem.* 77: 521-555.

Lal, R. (2005) World crop residues production and implications of its use as a biofuel. *Environ. Int.* 31: 575-584.

Lamond, A.I. (2002) Molecular biology of the cell, 4th edition. *Nature* 417: 383-383.

Lan, W., Lu, F.C., Regner, M., Zhu, Y.M., Rencoret, J., Ralph, S.A., et al. (2015) Tricin, a Flavonoid Monomer in Monocot Lignification. *Plant Physiol* 167: 1284-U1265.

Lancashire, P.D., Bleiholder, H., Vandenboom, T., Langeluddeke, P., Stauss, R., Weber, E., et al. (1991) A uniform decimal code for growth-stages of crops and weeds. *Ann Appl Biol* 119: 561-601.

Larsen, M.R., Thingholm, T.E., Jensen, O.N., Roepstorff, P. and Jorgensen, T.J.D. (2005) Highly selective enrichment of phosphorylated peptides from peptide mixtures using titanium dioxide microcolumns. *Molecular & Cellular Proteomics* 4: 873-886.

Lee, C., Teng, Q., Zhong, R., Yuan, Y. and Ye, Z.-H. (2014a) Functional roles of rice glycosyltransferase family GT43 in xylan biosynthesis. *Plant signaling & behavior* 9: e27809.

Lee, C., Teng, Q., Zhong, R., Yuan, Y. and Ye, Z.H. (2014b) Functional roles of rice glycosyltransferase family GT43 in xylan biosynthesis. *Plant Signal Behav* 9: e27809.

Lee, D.G., Houston, N.L., Stevenson, S.E., Ladics, G.S., McClain, S., Privalle, L., et al. (2010) Mass spectrometry analysis of soybean seed proteins: optimization of gel-free quantitative workflow. *Anal Methods-Uk* 2: 1577-1583.

Lee, R.C., Burton, R.A., Hrmova, M. and Fincher, G.B. (2001) Barley arabinoxylan arabinofuranohydrolases: purification, characterization and determination of primary structures from cDNA clones. *Biochem J* 356: 181-189.

Lee, R.C., Hrmova, M., Burton, R.A., Lahnstein, J. and Fincher, G.B. (2003) Bifunctional family 3 glycoside hydrolases from barley with alpha-L-arabinofuranosidase and beta-D-xylosidase activity - Characterization, primary structures, and COOH-terminal processing. *J Biol Chem* 278: 5377-5387.

Li, G.-Z., Vissers, J.P.C., Silva, J.C., Golick, D., Gorenstein, M.V. and Geromanos, S.J. (2009) Database searching and accounting of multiplexed precursor and product ion spectra from the data independent analysis of simple and complex peptide mixtures. *Proteomics* 9: 1696-1719.

Li, M., Heckwolf, M., Crowe, J.D., Williams, D.L., Magee, T.D., Kaeppler, S.M., et al. (2015) Cell-wall properties contributing to improved deconstruction by alkaline pre-treatment and enzymatic hydrolysis in diverse maize (Zea mays L.) lines. *J Exp Bot*.

Li, Q., Jain, M.R., Chen, W. and Li, H. (2013) A multidimensional approach to an in-depth proteomics analysis of transcriptional regulators in neuroblastoma cells. *J. Neurosci. Methods* 216: 118-127.

Li, R.Y. and Qu, R.D. (2011) High throughput Agrobacterium-mediated switchgrass transformation. *Biomass Bioenerg* 35: 1046-1054.

Li, Z.Y., Xia, J., Chen, Z., Yu, Y., Li, Q.F., Zhang, Y.C., et al. (2016) Large-scale rewiring of innate immunity circuitry and microRNA regulation during initial rice blast infection. *Sci Rep-Uk* 6: 25493.

Liljegren, S. (2010) Phloroglucinol stain for lignin. *Cold Spring Harb Protoc* 2010: pdb prot4954.

Lin, F., Manisseri, C., Fagerstrom, A., Peck, M.L., Vega-Sanchez, M.E., Williams, B., et al. (2016) Cell wall composition and candidate biosynthesis gene expression during rice development. *Plant Cell Physiol* 57: 2058-2075.

Lin, F., Waters, C.L., Mallinson, R.G., Lobban, L.L. and Bartley, L.E. (2015) Relationships between Biomass Composition and Liquid Products Formed via Pyrolysis. *Frontiers in Energy Research* 3.

Lionetti, V., Giancaspro, A., Fabri, E., Giove, S.L., Reem, N., Zabotina, O.A., et al. (2015) Cell wall traits as potential resources to improve resistance of durum wheat against Fusarium graminearum. *BMC Plant Biol.* 15.

Loque, D., Scheller, H.V. and Pauly, M. (2015) Engineering of plant cell walls for enhanced biofuel production. *Curr Opin Plant Biol* 25: 151-161.

Lynd, L.R., Laser, M.S., Brandsby, D., Dale, B.E., Davison, B., Hamilton, R., et al. (2008) How biotech can transform biofuels. *Nat Biotechnol* 26: 169-172.

Ma, J.F. and Yamaji, N. (2006) Silicon uptake and accumulation in higher plants. *Trends Plant Sci* 11: 392-397.

MacAdam, J.W. and Grabber, J.H. (2002) Relationship of growth cessation with the formation of diferulate cross-links and p-coumaroylated lignins in tall fescue leaf blades. *Planta* 215: 785-793.

Mach, J. (2015) Metabolic Crosstalk: Interactions between the Phenylpropanoid and Glucosinolate Pathways in Arabidopsis. *Plant Cell* 27: 1367-1367.

Majewska-Sawka, A. and Nothnagel, E.A. (2000) The multiple roles of arabinogalactan proteins in plant development. *Plant Physiol* 122: 3-9.

Malinovsky, F.G., Fangel, J.U. and Willats, W.G.T. (2014) The role of the cell wall in plant immunity. *Front Plant Sci* 5.

Martin, A.P., Brown, C.W., Nguyen, D.Q., Palmer, W.M., Furbank, R.T., Byrt, C.S., et al. (2017) Cell Wall Development in an Elongating Internode of Setaria. In *Genetics and Genomics of Setaria*. Edited by Doust, A. and Diao, X. pp. 211-238. Springer International Publishing, Cham.

Martin, A.P., Palmer, W.M., Brown, C., Abel, C., Lunn, J.E., Furbank, R.T., et al. (2016) A developing Setaria viridis internode: an experimental system for the study of biomass generation in a C4 model species. *Biotechnol Biofuels* 9: 45.

Mathew, S. and Abraham, T.E. (2004) Ferulic acid: An antioxidant found naturally in plant cell walls and feruloyl esterases involved in its release and their applications. *Crit. Rev. Biotechnol.* 24: 59-83.

Matschi, S., Werner, S., Schulze, W.X., Legen, J., Hilger, H.H. and Romeis, T. (2013) Function of calcium-dependent protein kinase CPK28 of Arabidopsis thaliana in plant stem elongation and vascular development. *Plant J* 73: 883-896.

Minic, Z. (2008) Physiological roles of plant glycoside hydrolases. *Planta* 227: 723-740.

Minic, Z., Rihouey, C., Do, C.T., Lerouge, P. and Jouanin, L. (2004) Purification and characterization of enzymes exhibiting beta-D-xylosidase activities in stem tissues of Arabidopsis. *Plant Physiol* 135: 867-878.

Mitchell, R.A., Dupree, P. and Shewry, P.R. (2007) A novel bioinformatics approach identifies candidate genes for the synthesis and feruloylation of arabinoxylan. *Plant Physiol* 144: 43-53.

Mitra, S.K., Chen, R.Q., Dhandaydham, M., Wang, X.F., Blackburn, R.K., Kota, U., et al. (2015) An autophosphorylation site database for leucine-rich repeat receptor-like kinases in Arabidopsis thaliana. *Plant J* 82: 1042-1060.

Mohnen, D. (2008) Pectin structure and biosynthesis. *Curr Opin Plant Biol* 11: 266-277.

Moldenhauer, K., Wilson, C.E., Counce, P. and Hardke, J. (2013) Rice growth and development. In *Arkansas rice production handbook*. Edited by Hardke, J.T. University of Arkansas Division of Agriculture Cooperative Extension Service, Little Rock, Arkansas.

Molinari, H.B., Pellny, T.K., Freeman, J., Shewry, P.R. and Mitchell, R.A. (2013) Grass cell wall feruloylation: distribution of bound ferulate and candidate gene expression in Brachypodium distachyon. *Front Plant Sci* 4: 50.

Moller, I., Sorensen, I., Bernal, A.J., Blaukopf, C., Lee, K., Obro, J., et al. (2007) High-throughput mapping of cell-wall polymers within and between plants using novel microarrays. *Plant J* 50: 1118-1128.

Morse, A.M., Whetten, R.W., Dubos, C. and Campbell, M.M. (2009) Post-translational modification of an R2R3-MYB transcription factor by a MAP Kinase during xylem development. *New Phytologist* 183: 1001-1013.

Mortimer, J.C., Miles, G.P., Brown, D.M., Zhang, Z., Segura, M.P., Weimar, T., et al. (2010) Absence of branches from xylan in Arabidopsis gux mutants reveals potential for simplification of lignocellulosic biomass. *Proc. Natl. Acad. Sci. U. S. A.* 107: 17409-17414.

Muraille, L., Aguié-Béghin, V., Chabbert, B. and Molinari, M. (2017) Bioinspired lignocellulosic films to understand the mechanical properties of lignified plant cell walls at nanoscale. *Scientific Reports* 7: 44065.

Mutwil, M., Klie, S., Tohge, T., Giorgi, F.M., Wilkins, O., Campbell, M.M., et al. (2011) PlaNet: Combined Sequence and Expression Comparisons across Plant Networks Derived from Seven Species. *Plant Cell* 23: 895-910.

Nam, K.H. and Li, J.M. (2002) BRI1/BAK1, a receptor kinase pair mediating brassinosteroid signaling. *Cell* 110: 203-212.

Nguema-Ona, E., Vicre-Gibouin, M., Gotte, M., Plancot, B., Lerouge, P., Bardor, M., et al. (2014) Cell wall O-glycoproteins and N-glycoproteins: aspects of biosynthesis and function. *Front Plant Sci* 5.

Ning, J., Zhang, B.C., Wang, N.L., Zhou, Y.H. and Xiong, L.Z. (2011) Increased Leaf Angle1, a Raf-Like MAPKKK That Interacts with a Nuclear Protein Family, Regulates Mechanical Tissue Formation in the Lamina Joint of Rice. *Plant Cell* 23: 4334-4347.

Nishitani, K. and Nevins, D.J. (1989) Enzymic Analysis of Feruloylated Arabinoxylans (Feraxan) Derived from Zea mays Cell Walls : II. Fractionation and Partial Characterization of Feraxan Fragments Dissociated by a Bacillus subtilis Enzyme (Feraxanase). *Plant Physiol* 91: 242-248.

Nozu, Y., Tsugita, A. and Kamijo, K. (2006) Proteomic analysis of rice leaf, stem and root tissues during growth course. *Proteomics* 6: 3665-3670.

Obel, N., Porchia, A.C. and Scheller, H.V. (2002) Dynamic changes in cell wall polysaccharides during wheat seedling development. *Phytochemistry* 60: 603-610.

Oh, M.H., Sun, J.D., Oh, D.H., Zielinski, R.E., Clouse, S.D. and Huber, S.C. (2011) Enhancing Arabidopsis leaf growth by engineering the BRASSINOSTEROID INSENSITIVE1 receptor kinase. *Plant Physiol* 157: 120-131.

Ohmiya, Y., Takeda, T., Nakamura, S., Sakai, F. and Hayashi, T. (1995) Purification and Properties of a Wall-Bound Endo-1,4-Beta-Glucanase from Suspension-Cultured Poplar Cells. *Plant Cell Physiol* 36: 607-614.

Oikawa, A., Joshi, H.J., Rennie, E.A., Ebert, B., Manisseri, C., Heazlewood, J.L., et al. (2010) An Integrative Approach to the Identification of Arabidopsis and Rice Genes Involved in Xylan and Secondary Wall Development. *PLoS One* 5.

Olson, S.N., Ritter, K., Rooney, W., Kemanian, A., McCarl, B.A., Zhang, Y.Q., et al. (2012) High biomass yield energy sorghum: developing a genetic model for C4 grass bioenergy crops. *Biofuel Bioprod Bior* 6: 640-655.

Olszewski, N.E., West, C.M., Sassi, S.O. and Hartweck, L.M. (2010) O-GlcNAc protein modification in plants: Evolution and function. *Biochimica et Biophysica Acta (BBA) - General Subjects* 1800: 49-56.

Ort, D.R., Zhu, X. and Melis, A. (2011) Optimizing Antenna Size to Maximize Photosynthetic Efficiency. *Plant Physiol* 155: 79-85.

Ouyang, S., Zhu, W., Hamilton, J., Lin, H., Campbell, M., Childs, K., et al. (2007) The TIGR Rice Genome Annotation Resource: Improvements and new features. *Nucleic Acids Res.* 35: D883-D887.

Palmer, N.A., Sattler, S.E., Saathoff, A.J., Funnell, D., Pedersen, J.F. and Sarath, G. (2008) Genetic background impacts soluble and cell wall-bound aromatics in brown midrib mutants of sorghum. *Planta* 229: 115-127.

Park, A.R., Cho, S.K., Yun, U.J., Jin, M.Y., Lee, S.H., Sachetto-Martins, G., et al. (2001) Interaction of the arabidopsis receptor protein kinase Wak1 with a glycine-rich protein, AtGRP-3. *J Biol Chem* 276: 26688-26693.

Park, Y.W., Tominaga, R., Sugiyama, J., Furuta, Y., Tanimoto, E., Samejima, M., et al. (2003) Enhancement of growth by expression of poplar cellulase in Arabidopsis thaliana. *Plant J* 33: 1099-1106.

Parry, D.W., Hodson, M.J. and Sangster, A.G. (1984) Some Recent Advances in Studies of Silicon in Higher-Plants. *Philos T Roy Soc B* 304: 537-&.

Passardi, F., Longet, D., Penel, C. and Dunand, C. (2004) The class III peroxidase multigenic in land plants family in rice and its evolution. *Phytochemistry* 65: 1879-1893.

Pattathil, S., Avci, U., Miller, J.S. and Hahn, M.G. (2012) Immunological approaches to plant cell wall and biomass characterization: Glycome Profiling. *Methods Mol Biol* 908: 61-72.

Paulsen, A.D., Mettler, M.S. and Dauenhauer, P.J. (2013) The Role of Sample Dimension and Temperature in Cellulose Pyrolysis. *Energ Fuel* 27: 2126-2134.

Pauly, M. and Keegstra, K. (2008) Cell-wall carbohydrates and their modification as a resource for biofuels. *Plant J* 54: 559-568.

Pauly, M. and Keegstra, K. (2010) Plant cell wall polymers as precursors for biofuels. *Curr Opin Plant Biol* 13: 305-312.

Pena, M.J., Kulkarni, A.R., Backe, J., Boyd, M., O'Neill, M.A. and York, W.S. (2016) Structural diversity of xylans in the cell walls of monocots. *Planta* 244: 589-606.

Penning, B.W., Hunter, C.T., 3rd, Tayengwa, R., Eveland, A.L., Dugard, C.K., Olek, A.T., et al. (2009) Genetic resources for maize cell wall biology. *Plant Physiol* 151: 1703-1728.

Petrik, D.L., Karlen, S.D., Cass, C.L., Padmakshan, D., Lu, F.C., Liu, S., et al. (2014) p-Coumaroyl-CoA:monolignol transferase (PMT) acts specifically in the lignin biosynthetic pathway in Brachypodium distachyon. *Plant J* 77: 713-726.

Porter, C.L., Jr. (1966) An Analysis of Variation Between Upland and Lowland Switchgrass, Panicum Virgatum L., in Central Oklahoma. *Ecology* v. 47: pp. 980-992-1966 v.1947 no.1966.

Rajasundaram, D., Selbig, J., Persson, S. and Klie, S. (2014) Co-ordination and divergence of cell-specific transcription and translation of genes in arabidopsis root cells. *Annals of Botany* 114: 1109-1123.

Ralph, J. (2010) Hydroxycinnamates in lignification. *Phytochem Rev* 9: 65-83.

Ralph, J., Grabber, J.H. and Hatfield, R.D. (1995) Lignin-Ferulate Cross-Links in Grasses - Active Incorporation of Ferulate Polysaccharide Esters into Ryegrass Lignins. *Carbohyd Res* 275: 167-178.

Rancour, D.M., Marita, J.M. and Hatfield, R.D. (2012) Cell wall composition throughout development for the model grass Brachypodium distachyon. *Front Plant Sci* 3: 266.

Rautengarten, C., Ebert, B., Herter, T., Petzold, C.J., Ishii, T., Mukhopadhyay, A., et al. (2011) The Interconversion of UDP-Arabinopyranose and UDP-Arabinofuranose Is Indispensable for Plant Development in Arabidopsis. *Plant Cell* 23: 1373-1390.

Redestig, H. and Costa, I.G. (2011) Detection and interpretation of metabolite-transcript coresponses using combined profiling data. *Bioinformatics* 27: I357-I365.

Rennie, E.A., Hansen, S.F., Baidoo, E.E.K., Hadi, M.Z., Keasling, J.D. and Scheller, H.V. (2012) Three Members of the Arabidopsis Glycosyltransferase Family 8 Are Xylan Glucuronosyltransferases. *Plant Physiol* 159: 1408-1417.

Rennie, E.A. and Scheller, H.V. (2014) Xylan biosynthesis. *Curr Opin Biotech* 26: 100-107.

Rezaei, P.S., Shafaghat, H. and Daud, W.M.A.W. (2016) Aromatic hydrocarbon production by catalytic pyrolysis of palm kernel shell waste using a bifunctional Fe/HBeta catalyst: effect of lignin-derived phenolics on zeolite deactivation. *Green Chem* 18: 1684-1693.

Richard, T.L. (2010) Challenges in Scaling Up Biofuels Infrastructure. *Science* 329: 793-796.

Righetti, P.G. and Boschetti, E. (2016) Global proteome analysis in plants by means of peptide libraries and applications. *J Proteomics* 143: 3-14.

Ryden, P., Sugimoto-Shirasu, K., Smith, A.C., Findlay, K., Reiter, W.-D. and McCann, M.C. (2003) Tensile Properties of Arabidopsis Cell Walls Depend on Both a Xyloglucan Cross-Linked Microfibrillar Network and Rhamnogalacturonan II-Borate Complexes. *Plant Physiol* 132: 1033-1040.

Salehi, H., Ransom, C.B., Oraby, H.F., Seddighi, Z. and Sticklen, M.B. (2005) Delay in flowering and increase in biomass of transgenic tobacco expressing the Arabidopsis floral repressor gene FLOWERING LOCUS C. *Journal of Plant Physiology* 162: 711-717.

San Clemente, H. and Jamet, E. (2015) WallProtDB, a database resource for plant cell wall proteomics. *Plant Methods* 11: 2.

Saravanan, R.S. and Rose, J.K.C. (2004) A critical evaluation of sample extraction techniques for enhanced proteomic analysis of recalcitrant plant tissues. *Proteomics* 4: 2522-2532.

Sasayama, D., Azuma, T. and Itoh, K. (2011) Involvement of cell wall-bound phenolic acids in decrease in cell wall susceptibility to expansins during the cessation of rapid growth in internodes of floating rice. *J Plant Physiol* 168: 121-127.

Sattler, S.E., Saballos, A., Xin, Z., Funnell-Harris, D.L., Vermerris, W. and Pedersen, J.F. (2014) Characterization of novel Sorghum brown midrib mutants from an EMS-mutagenized population. *G3 (Bethesda)* 4: 2115-2124.

Scheible, W.R. and Pauly, M. (2004) Glycosyltransferases and cell wall biosynthesis: novel players and insights. *Curr Opin Plant Biol* 7: 285-295.

Scheller, H.V. and Ulvskov, P. (2010) Hemicelluloses. *Annual Review of Plant Biology, Vol 61* 61: 263-289.

Schmer, M.R., Vogel, K.P., Mitchell, R.B., Dien, B.S., Jung, H.G. and Casler, M.D. (2012) Temporal and Spatial Variation in Switchgrass Biomass Composition and Theoretical Ethanol Yield. *Agron. J.* 104: 54-64.

Schuman, M.C. and Baldwin, I.T. (2016) The layers of plant responses to insect herbivores. *Annu. Rev. Entomol.* 61: 373-394.

Sharma, R., Cao, P., Jung, K.H., Sharma, M.K. and Ronald, P.C. (2013) Construction of a rice glycoside hydrolase phylogenomic database and identification of targets for biofuel research. *Front Plant Sci* 4: 330.

Shen, H., He, X.Z., Poovaiah, C.R., Wuddineh, W.A., Ma, J.Y., Mann, D.G.J., et al. (2012) Functional characterization of the switchgrass (Panicum virgatum) R2R3-MYB transcription factor PvMYB4 for improvement of lignocellulosic feedstocks. *New Phytologist* 193: 121-136.

Shen, H., Mazarei, M., Hisano, H., Escamilla-Trevino, L., Fu, C.X., Pu, Y.Q., et al. (2013) A genomics approach to deciphering lignin biosynthesis in switchgrass. *Plant Cell* 25: 4342-4361.

Sibout, R., Le Bris, P., Legee, F., Cezard, L., Renault, H. and Lapierre, C. (2016) Structural redesigning Arabidopsis lignins into alkali-soluble lignins through the expression of p-coumaroyl-CoA: monolignol transferase PMT. *Plant Physiol* 170: 1358-1366.

Silva, J.C., Gorenstein, M.V., Li, G.Z., Vissers, J.P.C. and Geromanos, S.J. (2006) Absolute quantification of proteins by LCMSE - A virtue of parallel MS acquisition. *Molecular & Cellular Proteomics* 5: 144-156.

Sinclair, T.R. and Sheehy, J.E. (1999) Erect leaves and photosynthesis in rice. *Science* 283: 1456-1457.

Slewinski, T.L. (2012) Non-structural carbohydrate partitioning in grass stems: a target to increase yield stability, stress tolerance, and biofuel production. *J Exp Bot* 63: 4647-4670.

Smith, P.K., Krohn, R.I., Hermanson, G.T., Mallia, A.K., Gartner, F.H., Provenzano, M.D., et al. (1985) Measurement of Protein Using Bicinchoninic Acid. *Anal. Biochem.* 150: 76-85.

Smith-Moritz, A.M., Hao, Z., Fernandez-Nino, S.G., Fangel, J.U., Verhertbruggen, Y., Holman, H.Y.N., et al. (2015) Structural characterization of a mixed-linkage glucan deficient mutant reveals alteration in cellulose microfibril orientation in rice coleoptile mesophyll cell walls. *Front Plant Sci* 6.

Sørensen, I., Domozych, D. and Willats, W.G.T. (2010) How Have Plant Cell Walls Evolved? *Plant Physiol* 153: 366-372.

Storey, J.D. and Tibshirani, R. (2003) Statistical significance for genomewide studies. *Proc. Natl. Acad. Sci. U. S. A.* 100: 9440-9445.

Suits, M.D., Zhu, Y., Taylor, E.J., Walton, J., Zechel, D.L., Gilbert, H.J., et al. (2010) Structure and kinetic investigation of Streptococcus pyogenes family GH38 alpha-mannosidase. *PLoS One* 5: e9006.

Sultana, A., Kumar, A. and Harfield, D. (2010) Development of agri-pellet production cost and optimum size. *Bioresource Technol* 101: 5609-5621.

Swaney, D.L., McAlister, G.C. and Coon, J.J. (2008) Decision tree-driven tandem mass spectrometry for shotgun proteomics. *Nat. Meth.* 5: 959-964.

Sylvester, A.W., Parker-Clark, V. and Murray, G.A. (2001) Leaf shape and anatomy as indicators of phase change in the grasses: Comparison of maize, rice, and bluegrass. *Am J Bot* 88: 2157-2167.

Taboada, A., Novo-Uzal, E., Flores, G., Loureda, M., Barcelo, A.R., Masa, A., et al. (2010) Digestibility of silages in relation to their hydroxycinnamic acid content and lignin composition. *J Sci Food Agr* 90: 1155-1162.

Tan, L., Eberhard, S., Pattathil, S., Warder, C., Glushka, J., Yuan, C.H., et al. (2013) An Arabidopsis Cell Wall Proteoglycan Consists of Pectin and Arabinoxylan Covalently Linked to an Arabinogalactan Protein. *Plant Cell* 25: 270-287.

Tanger, P., Field, J.L., Jahn, C.E., DeFoort, M.W. and Leach, J.E. (2013) Biomass for thermochemical conversion: targets and challenges. *Front Plant Sci* 4.

Tanger, P., Vega-Sánchez, M., Fleming, M., Tran, K., Singh, S., Abrahamson, J., et al. (2015) Cell Wall Composition and Bioenergy Potential of Rice Straw Tissues Are Influenced by Environment, Tissue Type, and Genotype. *Bioenergy research*: 1-18.

Taus, T., Kocher, T., Pichler, P., Paschke, C., Schmidt, A., Henrich, C., et al. (2011) Universal and confident phosphorylation site localization using phosphoRS. *J. Proteome Res.* 10: 5354-5362.

Tavares, E.Q.P., De Souza, A.P. and Buckeridge, M.S. (2015) How endogenous plant cell-wall degradation mechanisms can help achieve higher efficiency in saccharification of biomass. *J Exp Bot* 66: 4133-4143.

Taylor, N.G. (2007) Identification of cellulose synthase AtCesA7 (IRX3) in vivo phosphorylation sites - a potential role in regulating protein degradation. *Plant Mol. Biol.* 64: 161-171.

Taylor-Teeples, M., Lin, L., de Lucas, M., Turco, G., Toal, T.W., Gaudinier, A., et al. (2015) An Arabidopsis gene regulatory network for secondary cell wall synthesis. *Nature* 517: 571-U307.

Tilman, D., Socolow, R., Foley, J.A., Hill, J., Larson, E., Lynd, L., et al. (2009) Beneficial Biofuels-The Food, Energy, and Environment Trilemma. *Science* 325: 270-271.

Tor, M., Lotze, M.T. and Holton, N. (2009) Receptor-mediated signalling in plants: molecular patterns and programmes. *J Exp Bot* 60: 3645-3654.

Tronchet, M., BalaguÉ, C., Kroj, T., Jouanin, L. and Roby, D. (2010) Cinnamyl alcohol dehydrogenases-C and D, key enzymes in lignin biosynthesis, play an essential role in disease resistance in Arabidopsis. *Molecular Plant Pathology* 11: 83-92.

Underwood, W. (2012) The plant cell wall: a dynamic barrier against pathogen invasion. *Front Plant Sci* 3.

Urahara, T., Tsuchiya, K., Kotake, T., Tohno-oka, T., Komae, K., Kawada, N., et al. (2004) A b-(14)-xylosyltransferase involved in the synthesis of arabinoxylans in developing barley endosperms. *Physiol Plantarum* 122: 169-180.

Urbanowicz, B.R., Pena, M.J., Moniz, H.A., Moremen, K.W. and York, W.S. (2014) Two Arabidopsis proteins synthesize acetylated xylan in vitro. *Plant J* 80: 197-206.

US-DOE (2005) Biomass as Feedstock for a Bioenergy and Bioproducts Industry: e Technical Feasibility of a Billion-Ton Annual Supply

US-DOE (2011) U.S. Billion-Ton Update: Biomass Supply for a Bioenergy and Bioproducts Industry.

US-DOE (2015) Lignocellulosic Biomass for Advanced Biofuels and Bioproducts: Workshop Report.

US-DOE (2016) 2016 Billion-Ton Report: Advancing Domestic Resources for a Thriving Bioeconomy.

Uslu, A., Faaij, A.P.C. and Bergman, P.C.A. (2008) Pre-treatment technologies, and their effect on international bioenergy supply chain logistics. Techno-economic evaluation of torrefaction, fast pyrolysis and pelletisation. *Energy* 33: 1206-1223.

Van Acker, R., Vanholme, R., Storme, V., Mortimer, J.C., Dupree, P. and Boerjan, W. (2013) Lignin biosynthesis perturbations affect secondary cell wall composition and saccharification yield in Arabidopsis thaliana. *Biotechnology for biofuels* 6: 46.

van der Weijde, T., Kamei, C.L.A., Torres, A.F., Vermerris, W., Dolstra, O., Visser, R.G.F., et al. (2013) The potential of C4 grasses for cellulosic biofuel production. *Front Plant Sci* 4.

Vanholme, R., Demedts, B., Morreel, K., Ralph, J. and Boerjan, W. (2010a) Lignin Biosynthesis and Structure. *Plant Physiol* 153: 895-905.

Vanholme, R., Ralph, J., Akiyama, T., Lu, F.C., Pazo, J.R., Kim, H., et al. (2010b) Engineering traditional monolignols out of lignin by concomitant up-regulation of F5H1 and down-regulation of COMT in Arabidopsis. *Plant J* 64: 885-897.

Vanholme, R., Storme, V., Vanholme, B., Sundin, L., Christensen, J.H., Goeminne, G., et al. (2012) A systems biology view of responses to lignin biosynthesis perturbations in Arabidopsis. *Plant Cell* 24: 3506-3529.

Varhegyi, G., Antal, M.J., Jakab, E. and Szabo, P. (1997) Kinetic modeling of biomass pyrolysis. *J Anal Appl Pyrol* 42: 73-87.

Vega-Sanchez, M.E., Verhertbruggen, Y., Christensen, U., Chen, X., Sharma, V., Varanasi, P., et al. (2012) Loss of Cellulose Synthase-Like F6 function affects mixed-linkage glucan deposition, cell wall mechanical properties and defense responses in vegetative tissues of rice. *Plant Physiol*.

Verma, M., Godbout, S., Brar, S.K., Solomatnikova, O., Lemay, S.P. and Larouche, J.P. (2012) Biofuels Production from Biomass by Thermochemical Conversion Technologies. *International Journal of Chemical Engineering* 2012: 18.

Viajante, V. and Heinrichs, E.A. (1987) Plant-age effect of rice cultivar Ir46 on susceptibility to the yellow stem borer scirpophaga-incertulas (Walker) (Lepidoptera, Pyralidae). *Crop Prot.* 6: 33-37.

Vizcaino, J.A., Cote, R.G., Csordas, A., Dianes, J.A., Fabregat, A., Foster, J.M., et al. (2013) The proteomics identifications (PRIDE) database and associated tools: status in 2013. *Nucleic Acids Res.* 41: D1063-D1069.

Vogel, C. and Marcotte, E.M. (2012) Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat. Rev. Genet.* 13: 227-232.

Vogel, J. (2008) Unique aspects of the grass cell wall. *Curr Opin Plant Biol* 11: 301-307.

Waldron, K.W., Parker, M.L. and Smith, A.C. (2003) Plant Cell Walls and Food Quality. *Comprehensive Reviews in Food Science and Food Safety* 2: 128-146.

Walley, J.W., Sartor, R.C., Shen, Z.X., Schmitz, R.J., Wu, K.J., Urich, M.A., et al. (2016) Integration of omic networks in a developmental atlas of maize. *Science* 353: 814-818.

Wang, H., Chang-Wong, T., Tang, H.Y. and Speicher, D.W. (2010a) Comparison of extensive protein fractionation and repetitive LC-MS/MS analyses on depth of analysis for complex proteomes. *J. Proteome Res.* 9: 1032-1040.

Wang, H., Meng, J., Peng, X., Tang, X., Zhou, P., Xiang, J., et al. (2015a) Rice WRKY4 acts as a transcriptional activator mediating defense responses toward Rhizoctonia solani, the causing agent of rice sheath blight. *Plant Mol. Biol.* 89: 157-171.

Wang, H.Z. and Dixon, R.A. (2012) On-off switches for secondary cell wall biosynthesis. *Mol Plant* 5: 297-303.

Wang, J., Zhu, J., Huang, R. and Yang, Y. (2012) Investigation of cell wall composition related to stem lodging resistance in wheat (Triticum aestivum L.) by FTIR spectroscopy. *Plant Signal Behav* 7: 856-863.

Wang, J.P., Chuang, L., Loziuk, P.L., Chen, H., Lin, Y.C., Shi, R., et al. (2015b) Phosphorylation is an on/off switch for 5-hydroxyconiferaldehyde O-methyltransferase activity in poplar monolignol biosynthesis. *P Natl Acad Sci USA* 112: 8481-8486.

Wang, L., Guo, K., Li, Y., Tu, Y., Hu, H., Wang, B., et al. (2010b) Expression profiling and integrative analysis of the CESA/CSL superfamily in rice. *BMC Plant Biol* 10: 282.

Wang, Y.X., Yang, F., Gritsenko, M.A., Wang, Y.C., Clauss, T., Liu, T., et al. (2011) Reversed-phase chromatography with multiple fraction concatenation strategy for proteome profiling of human MCF10A cells. *Proteomics* 11: 2019-2026.

War, A.R., Paulraj, M.G., Ahmad, T., Buhroo, A.A., Hussain, B., Ignacimuthu, S., et al. (2012) Mechanisms of plant defense against insect herbivores. *Plant Signal Behav* 7: 1306-1320.

Warnasooriya, S.N. and Brutnell, T.P. (2014) Enhancing the productivity of grasses under high-density planting by engineering light responses: from model systems to feedstocks. *J Exp Bot* 65: 2825-2834.

Waters (2011) An overview of the principles of MSE, the engine that drives MS performance.

Werner, K., Pommer, L. and Brostrom, M. (2014) Thermal decomposition of hemicelluloses. *J Anal Appl Pyrol* 110: 130-137.

Wilkerson, C.G., Mansfield, S.D., Lu, F., Withers, S., Park, J.-Y., Karlen, S.D., et al. (2014) Monolignol Ferulate Transferase Introduces Chemically Labile Linkages into the Lignin Backbone. *Science* 344: 90-93.

Williams, W.P., Buckley, P.M. and Davis, F.M. (2000) Vegetative phase change in maize and its association with resistance to fall armyworm. *Maydica* 45: 215-219.

Willis, J.D., Smith, J.A., Mazarei, M., Zhang, J.Y., Turner, G.B., Decker, S.R., et al. (2016) Downregulation of a UDP-Arabinomutase Gene in Switchgrass (Panicum virgatum L.) Results in Increased Cell Wall Lignin While Reducing Arabinose-Glycans. *Front Plant Sci* 7.

Wilson, S.M., Ho, Y.Y., Lampugnani, E.R., Van de Meene, A.M., Bain, M.P., Bacic, A., et al. (2015) Determining the subcellular location of synthesis and assembly of the cell wall polysaccharide (1,3; 1,4)-beta-D-glucan in grasses. *Plant Cell* 27: 754-771.

Withers, S., Lu, F.C., Kim, H., Zhu, Y.M., Ralph, J. and Wilkerson, C.G. (2012) Identification of Grass-specific Enzyme That Acylates Monolignols with p-Coumarate. *J Biol Chem* 287: 8347-8355.

Wolf, S., Hematy, K. and Hofte, H. (2012) Growth Control and Cell Wall Signaling in Plants. *Annu Rev Plant Biol* 63: 381-407.

Wolf, S., Schmidt, S., Muller-Hannemann, M. and Neumann, S. (2010) In silico fragmentation for computer assisted identification of metabolite mass spectra. *BMC Bioinformatics* 11: 148.

Wullschleger, S.D., Davis, E.B., Borsuk, M.E., Gunderson, C.A. and Lynd, L.R. (2010) Biomass Production in Switchgrass across the United States: Database Description and Determinants of Yield. *Agron. J.* 102: 1158-1168.

Xu, B., Sathitsuksanoh, N., Tang, Y.H., Udvardi, M.K., Zhang, J.Y., Shen, Z.X., et al. (2012) Overexpression of AtLOV1 in Switchgrass Alters Plant Architecture, Lignin Content, and Flowering Time. *PLoS One* 7.

Xue, X. and Fry, S.C. (2012) Evolution of mixed-linkage (1 -> 3, 1 -> 4)-beta-D-glucan (MLG) and xyloglucan in Equisetum (horsetails) and other monilophytes. *Ann Bot* 109: 873-886.

Yamaji, N. and Ma, J.F. (2014) The node, a hub for mineral nutrient distribution in graminaceous plants. *Trends Plant Sci* 19: 556-563.

Yamaji, N., Mitatni, N. and Ma, J.F. (2008) A transporter regulating silicon distribution in rice shoots. *Plant Cell* 20: 1381-1389.

Yang, P.F., Liang, Y., Shen, S.H. and Kuang, T.Y. (2006) Proteome analysis of rice uppermost internodes at the milky stage. *Proteomics* 6: 3330-3338.

Yonekura-Sakakibara, K., Tohge, T., Matsuda, F., Nakabayashi, R., Takayama, H., Niida, R., et al. (2008) Comprehensive flavonol profiling and transcriptome coexpression analysis leading to decoding gene-metabolite correlations in Arabidopsis. *Plant Cell* 20: 2160-2176.

York, W.S. and O'Neill, M.A. (2008) Biochemical control of xylan biosynthesis - which end is up? *Curr Opin Plant Biol* 11: 258-265.

Yoshida, K. and Komae, K. (2006) A rice family 9 glycoside hydrolase isozyme with broad substrate specificity for hemicelluloses in type II cell walls. *Plant Cell Physiol* 47: 1541-1554.

Zeng, W., Jiang, N., Nadella, R., Killen, T.L., Nadella, V. and Faik, A. (2010) A glucurono(arabino)xylan synthase complex from wheat contains members of the

GT43, GT47, and GT75 families and functions cooperatively. *Plant Physiol* 154: 78-97.

Zhang, J., Choi, Y.S., Yoo, C.G., Kim, T.H., Brown, R.C. and Shanks, B.H. (2015) Cellulose-Hemicellulose and Cellulose-Lignin Interactions during Fast Pyrolysis. *Acs Sustain Chem Eng* 3: 293-301.

Zhang, K., Qian, Q., Huang, Z., Wang, Y., Li, M., Hong, L., et al. (2006) GOLD HULL AND INTERNODE2 encodes a primarily multifunctional cinnamyl-alcohol dehydrogenase in rice. *Plant Physiol* 140: 972-983.

Zhang, Q., Cheetamun, R., Dhugga, K.S., Rafalski, J.A., Tingey, S.V., Shirley, N.J., et al. (2014) Spatial gradients in cell wall composition and transcriptional profiles along elongating maize internodes. *BMC Plant Biol* 14: 27.

Zhao, K. and Bartley, L.E. (2014) Comparative genomic analysis of the R2R3 MYB secondary cell wall regulators of Arabidopsis, poplar, rice, maize, and switchgrass. *BMC Plant Biol* 14: 135.

Zhao, Q. and Dixon, R.A. (2014) Altering the Cell Wall and Its Impact on Plant Disease: From Forage to Bioenergy. *Annual Review of Phytopathology, Vol 52* 52: 69-91.

Zhao, S.Q., Xiang, J.J. and Xue, H.W. (2013) Studies on the rice LEAF INCLINATION1 (LC1), an IAA-amido synthetase, reveal the effects of auxin in leaf inclination control. *Mol Plant* 6: 174-187.

Zhong, R., Lee, C., Zhou, J., McCarthy, R.L. and Ye, Z.H. (2008) A battery of transcription factors involved in the regulation of secondary cell wall biosynthesis in Arabidopsis. *Plant Cell* 20: 2763-2782.

Zhong, R.Q., Lee, C., McCarthy, R.L., Reeves, C.K., Jones, E.G. and Ye, Z.H. (2011) Transcriptional Activation of Secondary Wall Biosynthesis by Rice and Maize NAC and MYB Transcription Factors. *Plant Cell Physiol* 52: 1856-1871.